Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Randomized Controlled Field Trials of Predictive Policing

G. O. Mohler, M. B. Short, Sean Malinowski, Mark Johnson, G. E. Tita, Andrea L. Bertozzi & P. J. Brantingham

To cite this article: G. O. Mohler, M. B. Short, Sean Malinowski, Mark Johnson, G. E. Tita, Andrea L. Bertozzi & P. J. Brantingham (2015) Randomized Controlled Field Trials of Predictive Policing, Journal of the American Statistical Association, 110:512, 1399-1411, DOI: 10.1080/01621459.2015.1077710

To link to this article: https://doi.org/10.1080/01621459.2015.1077710

Published online: 15 Jan 2016.

Randomized Controlled Field Trials of Predictive Policing

G. O. MOHLER, M. B. SHORT, Sean MALINOWSKI, Mark JOHNSON, G. E. TITA, Andrea L. BERTOZZI, and P. J. BRANTINGHAM

The concentration of police resources in stable crime hotspots has proven effective in reducing crime, but the extent to which police can disrupt dynamically changing crime hotspots is unknown. Police must be able to anticipate the future location of dynamic hotspots to disrupt them. Here we report results of two randomized controlled trials of near real-time epidemic-type aftershock sequence (ETAS) crime forecasting, one trial within three divisions of the Los Angeles Police Department and the other trial within two divisions of the Kent Police Department (United Kingdom). We investigate the extent to which (i) ETAS models of short-term crime risk outperform existing best practice of hotspot maps produced by dedicated crime analysts, (ii) police officers in the field can dynamically patrol predicted hotspots given limited resources, and (iii) crime can be reduced by predictive policing algorithms under realistic law enforcement resource constraints. While previous hotspot policing experiments fix treatment and control hotspots throughout the experimental period, we use a novel experimental design to allow treatment and control hotspots to change dynamically over the course of the experiment. Our results show that ETAS models predict 1.4–2.2 times as much crime compared to a dedicated crime analyst using existing criminal intelligence and hotspot mapping practice. Police patrols using ETAS forecasts led to an average 7.4% reduction in crime volume as a function of patrol time, whereas patrols based upon analyst predictions showed no significant effect. Dynamic police patrol in response to ETAS crime forecasts can disrupt opportunities for crime and lead to real crime reductions.

KEY WORDS: Crime; Experimental methods; Machine learning; Point processes; Policing dosage.

1. INTRODUCTION

Crime events arise out of interactions between local, place-based environmental conditions (Brantingham and Brantingham 1981; Weisburd 2008) and the situational decision making of offenders and victims (Matsueda, Kreager, and Huizinga 2006; Keizer 2008). In theory, police patrol can prevent crime on a day-to-day basis by altering or disrupting the environmental conditions suitable for crime (Sampson, Raudenbush, and Earls 1997; Weisburd and Eck 2004). However, experimental studies attempting to measure the effectiveness of different police patrol strategies show mixed results (Braga 2005; Farrington and Welsh 2005). Random patrol has a negligible impact (Kelling et al. 1974) because the risk of crime is not uniformly distributed in space or time (Sherman, Gartin, and Buerger 1989; Johnson et al. 2007a). Random patrol therefore allocates resources to locations that have little or no associated crime risk. Hotspot policing, by contrast, concentrates overwhelming resources in direct response to nonuniform crime patterns (Sherman and Weisburd 1995), leading to crime suppression not only at deployment locations, but also over a surrounding region through a diffusion of benefits (Guerette and Bowers 2009). Displacement of offenders appears to be incomplete in both theory (Short et al. 2010) and practice (Weisburd et al. 2006; Ratcliffe et al. 2011), indicating that hotspot policing produces a net reduction in crime.

However, the use of overwhelming resources obscures the relationship between patrol time and effect on crime. For example, in the Philadelphia foot patrol experiment (Ratcliffe et al. 2011), hotspots were patrolled for 16 hr a day, 5 days a week, by new recruits emerging from the police academy. Treatment hotspot areas exhibited 23% fewer crimes than control areas, but it is not known whether the same result could have been achieved at more modest deployment levels. At a more realistic scale, Sherman and Weisburd (1995) identified small clusters of addresses and sought targeted patrol in treatment clusters of 3 hr per day. While even this relatively low level of dosage was difficult to achieve, they did observe significant effects on both calls for service and crime and disorder in treatment hotspots. Given high call-to-service volumes and shrinking police budgets, 1–2 hr per day of patrol in determined hotspots is likely a more realistic number for many police agencies.

While studies such as the foot patrol experiment show that the rate at which patrol officers deter crime increases as the amount of patrol time increases, the goal of predictive policing is to increase the rate of crime deterrence under fixed resources by choosing those hotspots with the highest projected crime rates as the targets for patrol. In the Philadelphia foot patrol experiment,

G. O. Mohler is Assistant Professor, Department of Mathematics and Computer Science, Santa Clara University, Santa Clara, CA 95053 (E-mail: gmohler@scu.edu). M. B. Short is Assistant Professor, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 (E-mail: mbshort@math.gatech.edu). Sean Malinowski is Commander and Chief of Staff, Los Angeles Police Department, Los Angeles, CA 90012 (E-mail: sean.malinowski@lapd.lacity.org). Mark Johnson is Head of Analysis, Kent Police Service, United Kingdom (E-mail: mark.johnson@kent.pnn.police.uk). G. E. Tita is Professor, Department of Criminology, University of California, Irvine, Irvine, CA 92697 (E-mail: gtita@uci.edu). Andrea L. Bertozzi is Professor, Betsy Wood Knapp Chair for Innovation and Creativity, and Director of Applied Mathematics, Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095 (E-mail: bertozzi@math.ucla.edu). P. J. Brantingham, Department of Anthropology, University of California, Los Angeles, Los Angeles, CA 90095 (E-mail: pjb@anthro.ucla.edu). This work was supported in part by NSF grant DMS-0968309, AFOSR MURI grant FA9550-10-1-0569, ONR grant N000141010221, ARO grant W911NF1010472, ARO MURI grant W911NF-11-1-0332, ONR grant N000141210838, ARO MURI grant W911NF-11-1-0332, and NSF Grant SES-1343123. Special thanks to Greg Ridgeway and Omar Qazi. GM and PJB are co-founders of PredPol, a company offering predictive policing services to law enforcement agencies. GM, PJB, MBS, ALB, and GET hold stock in PredPol.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jasa.

Journal of the American Statistical Association
December 2015, Vol. 110, No. 512, Applications and Case Studies
DOI: 10.1080/01621459.2015.1077710

120 hotspots were selected by ranking all potential hotspots according to the crime volume over the previous 3 years, weighting previous years 1, 2, and 3 by 1.0, 0.5, and 0.25, respectively. Then each pair of hotspots in the sequential list were randomly partitioned to either treatment or control, yielding 60 treatment hotspots that were fixed throughout the experiment. If a more accurate ranking method were used to dynamically select 60 different hotspots each day, then our hypothesis is that a greater net reduction in crime would be possible under the same patrol resource constraints.

A number of algorithmic methods have been proposed for estimating crime hotspot risk. Multivariate models of crime hotspots include fixed environmental variables such as demographics (Wang and Brown 2012), income levels (Liu and Brown 2003), and distance to crime attractors (Liu and Brown 2003; Kennedy, Caplan, and Piza 2011; Wang and Brown 2012) to augment variables defined by crime incident volume. Risk factors that vary in space but not time help reduce variance and are effective for estimating long-term, chronic hotspots. Alternatively, short-term hotspots are estimated in dynamic hotspot mapping models of repeat victimization (Bowers, Johnson, and Pease 2004; Chainey, Tompson, and Uhlig 2008; Johnson et al. 2007b, 2009; Fielding and Jones 2012), where kernel density estimation is applied over a sliding window, typically on the order of several weeks, across the experimental period. In this article we use an epidemic-type aftershock sequence (ETAS) model (Mohler et al. 2011; Mohler 2014) that estimates the risk associated with both long-term hotspots and short-term models of near-repeat risk. Analogous to ETAS models of seismic activity (Marsan and Lengline 2008), a stationary point process representing fixed environmental risk is estimated via an expectation-maximization algorithm simultaneously with a short-term triggering kernel reflecting short-term, dynamic risk.

We conduct two experiments using the ETAS model, one in three divisions of the Los Angeles Police Department and one in two divisions of the Kent (United Kingdom) Police Department. In the experiments, the ETAS algorithm (treatment) was put head-to-head with hotspots maps produced each day and shift by dedicated crime analysts (control). We used silent tests where predictions were not deployed to the field to evaluate the predictive accuracy of the treatment method compared with control. In the three divisions within Los Angeles, we additionally conducted single-blind field trials where officers used either treatment or control predictions each day as the focus of their patrols. Because hotspot locations dynamically changed each day, we used an experimental design where days were randomly assigned to treatment or control. Thus, each day all officers within an entire division received either ETAS predictions or analyst predictions that were identical in appearance save for the hotspot locations. Our results show that ETAS models predict 1.4–2.2 times as much crime compared to a dedicated crime analyst using existing criminal intelligence and hotspot mapping practice. In the three divisions in Los Angeles, police patrols using ETAS forecasts led to an average 7.4% reduction in crime volume as a function of patrol time, whereas patrols based upon analyst predictions showed less than half of the treatment effect at a level that was not statistically significant.

The outline of the article is as follows. In Section 2, we provide details on our methodology. In Section 2.1, we outline our experimental design. In Section 2.2, we provide details on the ETAS algorithm and the cloud-based software architecture used in the experiment. In Section 2.3, we provide details on the methodology used by the crime analysts to produce the control hotspot maps. In Section 2.4, we provide details on the methodology used by the patrol officers in the field. In Section 2.5, we describe the data used by the ETAS algorithm and the analyst to produce hotspot maps. In Section 3, we present our findings. Section 3.1 presents results of the silent tests to evaluate model accuracy. In Section 3.2, we analyze police patrol activity in the field tests. In Section 3.3, we evaluate the impact of treatment and control predictions on crime rates in Los Angeles. Section 4 concludes with a discussion of the study in the broader context of crime forecasting and predictive policing.

2. METHODOLOGY

2.1 Design of Predictive Policing Field Experiments

Controlled experiments in social policy settings face unique constraints (Harrison and List 2004; Heckman 2008). The ideal counterfactual for a policing experiment would be to compare how a target subject would have responded in the absence of policing (i.e., a pure placebo). However, a placebo controlled design that would remove policing from the equation is not feasible for practical and ethical reasons. Rather, policing experiments are structured similarly to drug controlled trials (Kirk 1982), where a target policing strategy or tactic, designated as treatment, is compared with existing policing practice, designated as control. Randomization of subjects among treatment and control groups is the primary mechanism whereby a valid counterfactual is constructed. Subjects randomly assigned to a treatment group are exchangeable with those assigned to a control group. Thus, any effect of treatment would have been observed among control subjects had they been otherwise randomly assigned to the treatment condition. Conversely, no effect would have been observed among treatment subjects had they been randomly assigned to the control condition. Exchangeability is also essential in controlling for the vast number of unobserved factors. Indeed, it is both theoretically and practically impossible to observe and restrict all aspects of subject behavior. Randomization helps to ensure that treatment and control subjects are exchangeable with respect to these random effects.

Hotspot policing experiments have typically followed a partially random block design (Braga and Bond 2008; Ratcliffe et al. 2011). For example, Braga and Bond (2008) run a randomized complete block hotspot policing study that is of similar scale to the study here (individual hotspots on the order of 0.0115 sq miles, only 30% larger than the predictive policing boxes). Thirty-four hotspots were matched in 17 homogenous pairs and then randomly assigned to control and treatment conditions. However, their study is focused on the effects of targeting "shallow" problem solving at identified hotspots with routine patrol occurring in control hotspots. We are concerned with experimental comparison of different methods for identifying hotspots with equivalent policing tactics used in the different types of hotspots.

A different experimental design is required for evaluating predictive policing methods because the locations of hotspots are no longer fixed throughout the experiment. Instead, hotspot locations change based upon changes in estimated risk from the predictive model employed. Rather than dividing patrol areas into partially matched blocks, in our experimental design patrol "mission maps" were generated independently by a crime analyst (control) and ETAS algorithm (treatment) each day. Mission maps were identical in outward appearance, save for the exact placement of prediction boxes. Only after each set of predictions was generated was it randomly determined, using a Bernoulli random number generator, whether the control or treatment condition would be deployed to the field for the next 24-hr period (6 a.m.–6 a.m.). Importantly, command staff, supervisors, and patrol officers in each of the deployment areas were not aware of the distinction between treatment and control mission maps. Thus, days were randomly allocated to treatment or control and are considered exchangeable. Because analyst and ETAS prediction boxes are not constrained to be in the same place each day, but rather build up areas of exposure dynamically from day to day, it is feasible to measure the impact on crime at the aggregate scale of policing divisions. This is a unique feature of our experiment as previous hotspot policing studies by design only measure impact within the small area comprising the treatment and control hotspots.

Of course, choice among different experimental designs often involves tradeoffs. A random block design that compares treatment and control regions (Sherman and Weisburd 1995; Braga and Bond 2008; Ratcliffe et al. 2011) may ensure that treatment and control days are independent of one another, where there is no such guarantee for our design (but see further remarks regarding this in the Discussion). But this choice of a random block design comes at a cost whereby treatment and control regions cannot be guaranteed to be exchangeable. Indeed, the police officers assigned to and crime problems of one region are never strictly equivalent to those of another region no matter how much statistical matching one does. In keeping with the repeated measures type of experimental design used here, we are able to establish greater control over the type of confounds inherent to random block designs. Specifically, in our experiment the police officers involved are the same on both control and treatment days and they confront the same crime environment. We thus explore an experimental protocol with different strengths and weaknesses than those of random block design.

2.1.1 Silent Tests. We conducted silent experiments across two divisions of the Kent (United Kingdom) Police Department, Maidstone and Sevenoaks, from January 14, 2013, to April 7, 2013, and the Southwest Division of the Los Angeles Police Department (LAPD) from May 16, 2012, to January 10, 2013. Silent tests of predictive accuracy were conducted to control for biases that would be introduced by directed patrol in response to predictions. Each day both the ETAS algorithm and a crime analyst (one in Kent and one in Southwest) selected twenty 150 × 150 m prediction boxes for each of two 12-hr shifts in each division. Crime analysts in both settings followed their respective existing best practice in selecting predictions (see Section 2.3). The task set before each analyst was to place predictions to maximize the number of crimes captured over the subsequent 24-hr period. In Kent, predictions were not deployed to the field. We therefore compare head-to-head accuracy on each day of the silent test in Kent. In Southwest, it was randomly determined each day of the experiment whether to give ETAS or analyst predictions to the officers for directed patrol. We therefore compare the predictive accuracy of ETAS on days that ETAS predictions were not deployed to the field with the predictive accuracy of analyst predictions on days that analyst predictions were not deployed to the field. We discarded days on which the analyst was absent, yielding a total of 58 silent test days in Kent and 234 silent test days (117 ETAS and 117 analyst) in Southwest. To maintain independence of the ETAS and analyst box selection processes, mutual exclusivity of predictions was not enforced (see Section 3.3). In Maidstone and Sevenoaks Divisions in Kent, 13.5% and 14.1% of control and treatment boxes overlapped, respectively. In Southwest Division in Los Angeles, 9.1% of boxes overlapped.

2.1.2 Field Tests. To test whether predictive policing can impact crime, we conducted single-blind, randomized controlled field trials across three LAPD Divisions: Foothill from November 7, 2011, to April 27, 2012; North Hollywood from March 31, 2012, to September 14, 2012; and Southwest from May 16, 2012, to January 10, 2013. Following the finalization of daily mission maps it was randomly determined whether the ETAS or analyst missions would be deployed to the field. LAPD command staff, supervisors, and patrol officers were not aware of the distinction between control and treatment conditions, consistent with a single-blind experimental design (Kirk 1982). Patrol officers were directed to use available time to "get in the box" and police what they saw. Thus, control and treatment conditions contrast directed patrol patterns in space and time, not differences in field tactics (see Section 2.4 for further details on police activity). We tracked the amount of available time spent in prediction boxes using the existing in-car call logging system. As with silent accuracy testing, field-deployed missions consisted of twenty 150 × 150 m prediction boxes for each 12-hr shift per division. Prediction boxes were therefore free to vary in their locations each shift. We discarded days the analyst was absent, yielding samples of 124, 152, and 234 test days in each division, respectively. Since each day was randomly assigned to control or treatment, variation in both policing tactics and patrol time was independent of experimental condition, providing an opportunity to measure the impact of different patrol levels (Koper 1995; Telep, Mitchell, and Weisburd 2014). A total of 62 control days and 62 treatment days in Foothill, 82 control days and 70 treatment days in North Hollywood, and 117 control and 117 treatment days in Southwest were randomly assigned during the trial (Total = 510 days).

2.2 Epidemic-Type Aftershock Sequence Model for Crime Prediction

Building on a foundation of reaction-diffusion models of crime (Short et al. 2010), we treat the dynamic occurrence of crime as a continuous time, discrete space ETAS point process (Marsan and Lengline 2008; Mohler et al. 2011; Mohler 2014). Policing areas were first discretized into 150 × 150 m square boxes. The conditional intensity, or probabilistic rate $\lambda_n(t)$ of events in box n at time t was determined by

$$\lambda_n(t) = \mu_n + \sum_{t_n^i < t} \theta \omega e^{-\omega(t - t_n^i)}, \tag{1}$$

where t_n^i are the times of events in box n in the history of the process. The ETAS model has two components, one modeling place-based environmental conditions that are constant in time and the other modeling dynamic changes in risk. Rather than modeling fixed environmental characteristics of a hotspot explicitly using census data or locations of crime attractors, long-term hotspots are estimated from the events themselves. In particular, the background rate μ is a nonparametric histogram estimate of a stationary Poisson process (Marsan and Lengline 2008). If over the past 365 days a grid cell has a high crime volume, the estimate of μ will be large in that grid cell. The size of the grid cells on which μ is defined can be estimated by maximum likelihood and in general the optimum size of the grid cell will decrease with increasing data. However, for a fixed area flagged for patrol, a greater number of small hotspots are more difficult to patrol than a small number of large hotspots. The 150×150 m hotspots were chosen in this study to be the size of a city block in Foothill and were then held constant across all of the experimental regions. The number of days of data used as input to the ETAS model, 365 days, was also chosen subjectively, though is consistent with other hotspot policing studies that use 1–2 years of data to select hotspots.

The second component of the ETAS model is the triggering kernel $\theta\omega e^{-\omega t}$ that models "near-repeat" or "contagion" effects in crime data. The exponential decay causes grid cells containing recent crime events to have a higher intensity than grid cells with fewer recent events and the same background rate. The main difference between the ETAS model and prospective hotspot maps (Bowers, Johnson, and Pease 2004) that model near-repeat effects is the introduction of the background rate μ. Whereas prospective hotspot maps only take into account short-term hotspot dynamics, the ETAS model estimates both long-term and short-term hotspots and systematically estimates the relative contribution to risk of each via expectation-maximization (EM)(Mohler et al. 2011; Mohler 2014). Given an initial guess for the parameters θ, μ, and ω, the EM algorithm is applied iteratively until convergence by alternating between the following two steps:

E-step

$$p_n^{ij} = \frac{\theta\omega e^{-\omega(t_n^j - t_n^i)}}{\lambda_n(t_n^j)}, \tag{2}$$

$$p_n^{j} = \frac{\mu_n}{\lambda_n(t_n^j)}, \tag{3}$$

M-step

$$\omega = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_{i<j} p_n^{ij}(t_n^j - t_n^i)}, \tag{4}$$

$$\theta = \frac{\sum_n \sum_{i<j} p_n^{ij}}{\sum_n \sum_j 1}, \tag{5}$$

$$\mu = \frac{\sum_n \sum_j p_n^{j}}{T}, \tag{6}$$

where T is the length of the time window of observation.

The EM algorithm can be intuitively understood by viewing the ETAS model as a branching process (Mohler et al. 2011). First-generation events occur according to a Poisson process with constant rate μ. Events (from all generations) each give birth to N direct offspring events, where N is a Poisson random variable with parameter θ. As events occur, the rate of crime increases locally in space, leading to a contagious sequence of "aftershock" crimes (Mohler et al. 2011) that eventually dies out on its own, or is interrupted by police intervention; the former occurs naturally so long as $\theta < 1$, while the latter is unaccounted for by the model. In the E-step, the probability that event j is a direct offspring of event i is estimated, along with the probability that the event was generated by the Poisson process μ. Given the probabilistic estimate of the branching structure, the complete data log-likelihood is then maximized in the M-step, providing an estimate of the model parameters. For a detailed treatment of the EM algorithm in the context of ETAS see Veen and Schoenberg (2008) or Lewis and Mohler (2011) where the EM algorithm is shown to be equivalent to projected gradient ascent optimization on the log-likelihood.

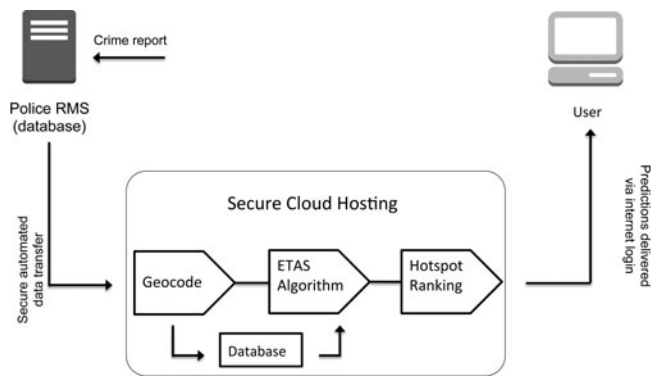

Figure 1. The ETAS model was implemented as a fully automated cloud-based machine learning system. Java software was installed on the police RMS server that encrypted and sent crime report data (address, date, time, crime type) once every hour to cloud-based servers where the data were geocoded using Google API and stored for later use. Once a day at 4 a.m. the ETAS parameters were reestimated using the previous 365 days of crime data (up to the latest records) and boxes were ranked by intensity λ_n. Patrol commanders logged on via a web interface prior to each patrol shift and printed out a report that was then distributed to officers. The report contained a map of hotspot locations as well as a list of nearest cross-streets of the hotspot locations.

An ETAS crime forecast consists of the top 20 boxes in rank order displayed on a map and available for directed police patrol. The algorithm was implemented as a fully automated cloud-based software system (see Figure 1). For each new crime report of the target crime types entered into the police records management system, the date, time, crime type, and address were automatically sent to secure, cloud-based data center servers. Events were geocoded in the cloud servers using Google API and stored for later use. Each day at 4 a.m. local time the ETAS algorithm was run and each 150×150 m cell in each division and shift was assigned a probability score by the algorithm. The top 20 cells were then displayed on a Google map that could be accessed by a user via a secure internet connection. The software also allowed the analyst on control days to place 20 boxes (identical in size and appearance to ETAS boxes) on the Google map instead of ETAS-generated boxes. Reports were printed out at roll call and distributed to the officers.

The ETAS model was automated to produce treatment missions using the most recent crime data available in the live records management system. Predictions were therefore generated in near real-time given natural lags in recorded crime data. Missions were set daily at 5 a.m. for two sequential shifts from 6 a.m. to 6 p.m. and 6 p.m. to 6 a.m. In Foothill, mission predictions were made for the 24-hr period starting at 4 p.m. The number of prediction boxes and prediction time windows in each policing division were identical to those generated by analysts.

2.3 Analyst Control Predictions

We compare predictions derived from ETAS to the best practices of trained crime analysts. Analysts were tasked to place a fixed number of 150 × 150 m boxes within their operational environment in specified time windows. The goal of box placement was to identify a small set of locations where the analysts expected crime to be most likely to occur. In Kent, one analyst was responsible for generating 80 prediction boxes each day, evenly divided among two policing divisions (Maidstone and Sevenoaks) and two shifts (7 a.m.–7 p.m. and 7 p.m.–7 a.m.). In Los Angeles, three crime analysts acting independently were responsible for placing prediction boxes for Foothill, North Hollywood, and Southwest Divisions. In Foothill Division, 20 predictions were generated for each 24-hr period beginning at 4 p.m. each day. In both North Hollywood and Southwest Divisions, 40 prediction boxes were generated each day evenly divided between two shifts (6 a.m.–6 p.m. and 6 p.m.–6 a.m.).

Crime analysts were free to use any information and methods at their disposal to generate predictions. Not surprisingly, therefore, the approaches taken to crime prediction differed substantially between Kent Police and Los Angeles Police Department analysts. Kent Police adhere to an intelligence-led policing approach (Ratcliffe 2012) as specified by the U.K. National Intelligence Model (National Centre for Policing Excellence 2005). The goal of the intelligence-led approach is to fuse detailed individual-level information on chronic or prolific offenders, who are responsible for a large fraction of the crime, with information on recent criminal incidents, known high-risk issues, and priority locations (i.e., hotspots), and use this information to guide resource allocation. The intelligence-led approach is explicit in its view that criminal intelligence helps police avoid excessive responses to random events. When tasked with placing prediction boxes, the Kent analyst relied primarily on information about the activity anchor points (e.g., home and work locations) of chronic offenders and their known associates. Information about recent crime events and their distribution in space and time was evaluated as relevant in relation to criminal intelligence. For example, a residential burglary that occurs in proximity to a place frequented by a known prolific offender is much more important in driving prediction placement than a burglary occurring in a location with no known connection to a prolific offender. Analysts in England and Wales consider crime mapping and analysis in support of intelligence-led policing a core function, though only about half of analysts from police forces report that such activities occupy most of their time (Weir and Bangs 2007).

The emphasis in Los Angeles was reversed. The Los Angeles Police Department follows a COMPSTAT (Walsh 2001) policing model focused on the analysis of 7-day crime maps supplemented with ad hoc street-level intelligence. The LAPD analysts emphasized small clusters of recent crimes as signaling emerging problems in need of short-term response to interrupt the formation of crime hotspots. Box placement therefore reflected the colloquial mantra about COMPSTAT as being "cops-on-the-dots." Here the analyst approach was "box-on-the-dots." Hotspot mapping is in widespread use within U.S. policing (Chainey and Ratcliffe 2005). A substantial majority of police forces serving populations of 100,000 people or more make use of hotspot mapping for both evaluation of past performance and tactical resource allocation (O'Shea et al. 2003; Friedmann, Rosenfeld, and Borissova 2010; Reaves 2010).

In both Kent and Los Angeles an easy-to-use web interface was developed to allow analysts to choose prediction locations by clicking on a Google Map. While the physical task of placing predictions could be completed in a matter of minutes, in practice analysts spent considerable time engaged in analysis of crime and intelligence data before committing to predictions. In Kent, the analyst spent between 3 hr and 4 hr each day researching chronic offender information and ongoing crime problems before placing their 80 total prediction boxes in the operational environment. In Los Angeles, the three analysts each spent 1.5–2 hr per day inspecting crime maps and recent crime reports before placing their 40 total prediction boxes (20 in Foothill).

2.4 Policing Activity Under Experimental Conditions

Police officers and patrol teams develop preferences that govern where they like to patrol and the types of policing tactics they like to use. For example, certain teams may prefer policing misdemeanor crimes, others may prefer to interdict chronic violent or property offenders, while still others may prefer non-crime community interactions. It is reasonable to suppose that some of these policing tactics will have more of an impact on suppressing crime than others, though there is continuous, vigorous debate surrounding policing tactical effectiveness. We are agnostic about the benefits of different tactical choices of officers in the field. In any case, LAPD command staff in each of the three test divisions did not mandate specific tactical actions. Rather, patrol officers were encouraged to use available time to "get in the box" and use their discretion to select appropriate field tactics upon entering prediction locations. In this approach we follow closely Sherman and Weisburd (1995) in that we neither restrict how police deal with individual hotspots, nor do we restrict the amount of time they engage the problem. Nevertheless, we were concerned with potential bias arising from any nonrandom distribution of police tactical preferences across control and treatment groups.

The experiment was designed to help guard against such biases. Since each day is randomly assigned to control or treatment condition at the start of the day, over time each experimental condition is expected to receive equal exposure to the range of policing tactics present among the patrol teams. For example, an officer or patrol team that prefers so-called "broken windows" policing—enforcing against misdemeanor infractions with the notion that it will disrupt felonies by the same individuals—will be provided equal opportunity to engage those tactics on control and treatment days (both inside and outside of prediction boxes). Moreover, because the outward appearance of control and treatment missions was identical, except for the exact placement of boxes, there were no cues that would encourage patrol teams to systematically adjust their tactics to be different on control or treatment days.

An ideal test of successful randomization of officer or patrol team tactical preferences would involve collection of detailed activity logs. This approach was not feasible given the scale and duration of the present experiment. Rather, we used officer time on mission in control and treatment prediction boxes as our primary measure of officer activity. Officers generated a time-stamped call event on their in-car mobile data terminal upon entering and leaving any mission box. We refer to the total amount of time spent within prediction boxes as officer time on mission and each incident of an officer spending time within a box as a mission stop.

2.5 Crime Types and Predictive Accuracy

The experiment focused on burglary, car theft, burglary-theft from vehicle, criminal damage, violence against the person (including sexual offences), and robbery as target crime types in Kent. These crimes together comprise 56% of the crime volume recorded by Kent. Burglary, car theft, and burglary-theft from vehicle were the target crime types in Los Angeles, which together comprise 55% of the reported crime volume. All of these crime types are known to be underreported (Mosher, Miethe, and Hart 2010) and therefore comparisons are made between control and treatment conditions subject to the same underreporting constraint. Reported crimes were tabulated according to whether they occurred strictly inside or outside active control and treatment boxes. A crime occurring strictly inside an active prediction box is considered a successful prediction. We do not consider here “near misses” or crimes that occurred within some small distance of an active prediction box. An active prediction was defined as a box generated before the start of a patrol shift and held constant in that location for the duration of the shift. Active predictions were replaced at the end of each shift by new active boxes. In the case of silent accuracy testing, active predictions were not made available to patrol officers. For deployed testing, patrol officers had access to maps of active prediction boxes during their patrol hours.

3. RESULTS

3.1 Accuracy of ETAS Versus Analyst Predictions

We first analyze predictive accuracy. In Table 1, we count the number of crimes occurring in active prediction boxes under silent experimental conditions. The analyst in Kent predicted 6.8% (Maidstone) and 4.0% (Sevenoaks) of crimes successfully compared to 9.8% and 6.8% by the ETAS model, a factor of 1.4 and 1.7 improvement in accuracy significant at the $p = 0.03$ and $p = 0.04$ level, respectively. In Southwest Division, over the 117 days in which control conditions were not deployed to the field, the analyst successfully predicted 2.1% of crimes. Over the 117 days in which treatment conditions were not deployed to the field, the ETAS model successfully predicted 4.7% of crimes. Treatment conditions thus yielded a predictive accuracy 2.2 times greater than control in the absence of police patrol effects ($p = 0.0012$).

The difference in the accuracy between analyst and ETAS predictions reflects advantages that algorithmic approaches have in characterizing dynamic spatio-temporal patterns (Mohler et al. 2011; Wang and Brown 2012; Mohler 2014). In LAPD Southwest Division, where the emphasis is on crime mapping, analyst predictions closely track recent crimes across the operational environment (Figures 2(a) and 2(b)). This indicates a tendency to judge overnight crime as more salient than other criteria in predicting future crime. In Kent Maidstone Division, the emphasis is on recent information about known prolific or chronic offenders, who can be responsible for a large proportion of crime (Kennedy 1996). The intelligence-led approach used in Kent produced more clustering in predictions over time (Figure 2(e)), reflecting an explicit rationale to not chase crimes that cannot be spatially associated with known chronic offenders (National Centre for Policing Excellence 2005) (Figure 2(d)).

Table 1. Successfully predicted crimes under nondeployed conditions.

	ETAS				Analyst					
	Success	Total	Rate	PAI	Success	Total	Rate	PAI	Boost	*P*-value
Maidstone	60	615	9.8%	85.2	42	615	6.8%	59.6	1.4	0.0314
Sevenoaks	39	576	6.8%	55.7	23	576	4.0%	32.8	1.7	0.0409
Southwest	46	986	4.7%	3.5	20	933	2.1%	1.6	2.2	0.0184

NOTE: Successful predictions are defined as crimes occurring strictly within an active 150 × 150 m box. Total crime in Maidstone and Sevenoaks, Kent, includes all burglary, car theft, theft from vehicle, and criminal damage occurring during 12-hr shifts where both analyst and ETAS predictions were generated. Total crime in Southwest, Los Angeles, is all burglary, car theft, and theft from vehicle occurring on corresponding days where treatment (ETAS) or control (analyst) predictions were not deployed to officers. PAI is the predictive accuracy index, an area-standardized measure of accuracy. Boost is the increase in predictive accuracy associated with treatment. *P*-value corresponds to a one-tailed, two-sample proportions test for the difference in mean predictive accuracy between treatment and analyst. Predictions are only calculated on shifts with exactly 20 prediction boxes. Random predictions correspond to an accuracy of 1.0% in Southwest and 0.1% in Maidstone and Sevenoaks.

Deployment of prediction boxes to the field and delivery of policing dosage to those boxes is expected to suppress some fraction of crime in those locations. Predictive accuracy should therefore decline in response to directed patrol. Table 2 shows that crimes did occur in treatment and control boxes deployed to the field. In LAPD Southwest Division, where we have a silent test of accuracy for comparison, predictive accuracy falls from 4.7% to 3.9%, a decline of −17% overall. Importantly, predictive accuracy for analyst control conditions remains effectively unchanged from the silent test (2.1%) to deployed settings (2.2%), consistent with the observation that policing dosage has limited effect under control conditions (see below). Table 2 also presents accuracies for treatment and control boxes under deployed conditions in LAPD Foothill and North Hollywood Divisions. Silent tests of accuracy were not conducted in these policing areas and therefore similar comparisons are not possible at present.

It is easy to predict a large fraction of crime if one designates a large area as being at risk for crime (Chainey, Tompson, and Uhlig 2008). In the experimental design used here, by contrast, we only place a tiny fraction of the total land area under prediction at any one time. LAPD Foothill, North Hollywood, and Southwest Divisions total 119.48, 64.87, and 33.96 km^2 in area, respectively. Only 0.38%, 0.69%, and 1.32% of the total areas, respectively, were under prediction. In Kent, the fraction of land area under prediction was even smaller. Sevenoaks is 370 km^2 in area and prediction boxes covered only 0.12% of the total at any

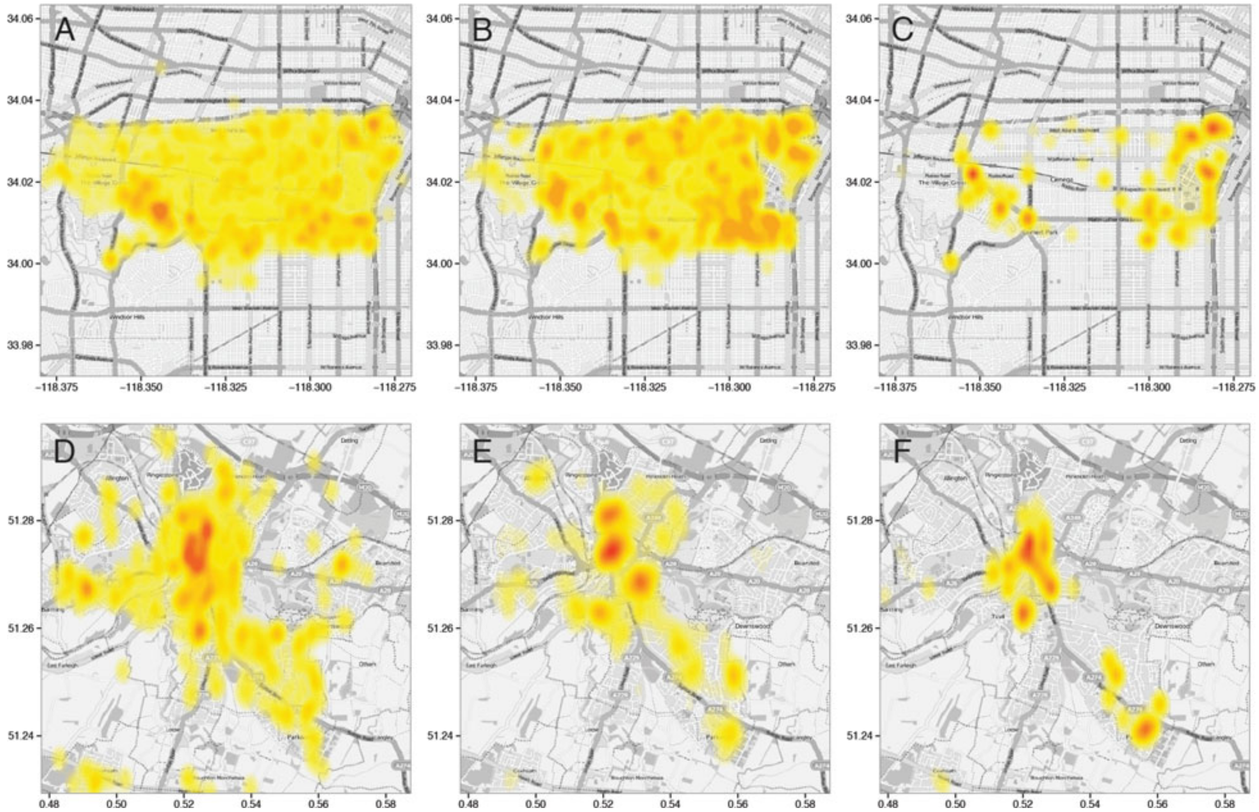

Figure 2. Crime is widely distributed in both (a) LAPD Southwest Division and (d) Kent Police Maidstone District. Intense crime hotspots are more prominent in Maidstone. (b) Control prediction boxes placed by the Southwest Division crime analyst tracked day-to-day point locations of crime leading to wide dispersion of boxes with numerous intermediate intensity prediction hotspots. (c) Treatment boxes placed by the ETAS model in Southwest are densely clustered and distinctive from both crime and analyst prediction box distributions. (e) Control prediction boxes placed by the Maidstone analyst are clustered compared with the Maidstone crime distribution and contrast with the dispersed predictions of the Southwest Division analyst, reflecting the greater focus on the activity locations of known offenders. (f) ETAS predictions in Maidstone are even more tightly clustered relative to crime and analyst prediction boxes. Shown in (a) is a random sample of 1000 crime locations in LAPD Southwest Division and in (d) all 485 recorded events in Maidstone, Kent. Shown in (b)–(c) and (e)–(f) are random samples of approximately 1000 prediction boxes representing ≈15% of control and treatment predictions in Southwest and ≈10.3% of control and treatment predictions in Maidstone. Maps show kernel density estimates with constant bandwidth across all cases.

one time. Maidstone is 393 km^2 in areas and prediction boxes covered only 0.11% of the total land area at any one time. The fraction of crime predicted relative to the fraction of land area under prediction is very large. In Maidstone, for example, 9.8% of crime is accurately predicted in just 0.11% of the land area, representing area-standardized predictive accuracy of 85.2.

Table 2. Successfully predicted crimes under deployed conditions

	ETAS				Analyst					
	Success	Total	Rate	PAI	Success	Total	Rate	PAI	Boost	P-value
Foothill	22	346	6.4%	16.9	11	347	3.2%	8.4	2.0	0.0244
N. Hollywood	21	611	3.4%	4.9	12	732	1.6%	2.4	2.1	0.0170
Southwest	38	981	3.9%	2.9	21	936	2.2%	1.7	1.7	0.0194
Total	81	1938	4.2%	6.8	44	2015	2.2%	3.5	1.9	0.0002

NOTE: Successful predictions defined as crime occurring strictly within an active 150 × 150 m box. Total crime is all burglary, car theft, and theft from vehicle occurring on corresponding treatment (ETAS) or control (analyst) days. PAI is the predictive accuracy index, an area-standardized measure of accuracy. Boost is the increase in predictive accuracy associated with treatment. P-value corresponds to a one-tailed, two-sample proportions test for the difference in mean predictive accuracy between treatment and analyst. Predictions are only calculated on shifts with exactly 20 prediction boxes. This precludes some shifts in Southwest in which the analyst created fewer than 20 boxes, explaining the discrepancy between analyst and ETAS total crime numbers in deployed versus nondeployed conditions.

3.2 Crime Analyst Methodology

The goal of this study was to compare a fully automated statistical algorithm (ETAS) for determining patrol hotspots to existing best practice in the Los Angeles and Kent police departments. Rather than attempting to control for analyst methodology or documenting analyst methodology through surveys, we perform a post-experiment analysis of the hotspots selected by the analysts during the trial. In Table 3, we compare analyst success rate with hotspot predictions corresponding to N-day hotspot maps. In Table 4, we compare analyst and hotspot prediction locations by calculating the number of identical hotspots for each day of the experiment aggregated over the experimental period. The hotspot maps are determined by ranking grid cells by the number of crimes occurring in the past N days. The analyst in Maidstone significantly outperforms 3-day hotspot maps in predicting crime on a shift-by-shift basis. The success rate of analysts in Sevenoaks and Southwest Division is statistically indistinguishable from 3-day hotspot maps. All three analysts are statistically indistinguishable from 7-day hotspot maps in predicting crime. By comparison, ETAS substantially outperforms both 3-day and 7-day hotspot maps. Consistent with the differ-

Table 3. Successfully predicted crimes for N-day hotspot maps.

	3-Day						7-Day					
	Success	Total	Rate	PAI	Boost	*P*-value	Success	Total	Rate	PAI	Boost	*P*-value
Maidstone	18	615	2.9%	24.1	2.3	0.0007	38	615	6.2%	50.8	1.1	0.3219
Sevenoaks	17	576	3.0%	25.8	1.4	0.1671	20	576	3.5%	30.3	1.2	0.3205
Southwest	17	933	1.8%	1.4	1.2	0.3092	27	933	2.9%	2.2	0.7	0.8495

NOTE: Successful predictions defined as crime occurring strictly within an active 150 × 150 m box. Total crime is all burglary, car theft, and theft from vehicle occurring on corresponding control (analyst) days. PAI is the predictive accuracy index, an area-standardized measure of accuracy. Boost is the increase in predictive accuracy associated with control compared with 3-day and 7-day hotspot maps. *P*-value corresponds to a one-tailed, two-sample proportions test for the difference in mean predictive accuracy between treatment and analyst. Predictions are only calculated on shifts with exactly 20 prediction boxes.

ence between ETAS and analysts, ETAS doubles the amount of crime predicted relative to hotspot maps with short measurement time windows.

In spite of the statistical similarity in performance in predicting crime, analyst predictions do deviate from strict hotspot mapping a fraction of the time. We observe in Table 4 that 13%–40% of analyst hotspots are identical to either 3-, 7-, 14-, 21-, or 28-day hotspot maps, depending on the analyst and division. In Table 5, we observe that analysts also place a significant number of hotspots to account for near-repeat victimization (Farrell and Pease 1993; Townsley, Homel, and Chaseling 2003). The percentage of analyst hotspot maps in agreement or adjacent to N-day hotspot maps ranges from 47% to 67%. There are several possible explanations for the remaining analyst hotspots not identical or adjacent to hotspots determined through hotspot mapping. One explanation is that analysts use a more complex function (compared to simple crime counts) when determining hotspot rankings. Another is that the remaining hotspots are determined through street level intelligence, for example location of parolees or known crime attractors. A third explanation is that constraints on working memory influence the selection of 20 hotspots for each shift. In Table 6, we observe some evidence for the third explanation as 7%–20% of analyst hotspots contained no crime throughout the course of the experiments.

3.3 Prediction Similarity Through Time

A distinctive feature of the present experiment is that crime prediction locations are free to vary on a shift-by-shift basis. In spite of this freedom, a fraction of prediction locations occurs repeatedly over the course of the experiment. Here we investigate the degree to which treatment and control predictions fall in the same locations on subsequent days. The method we use is analogous to a Hamming distance between prediction sets. We first choose a focal patrol shift and record the locations of all prediction boxes during that shift. We then look at corresponding patrol shifts on each subsequent day N and score the fraction of predictions falling in the same locations as boxes on the focal day. For example, if all of the 20 treatment predictions for the daytime shift on February 1, 2013, were repeated during the daytime shift on February 2, 2013, then the fraction of boxes repeated on day $N = 1$ is 1. If 10 of the prediction from February 1, 2013, were repeated on February 11, then the fraction of boxes repeated on day $N = 10$ is 0.5. Note that static hotspot policing experiments establish treatment and control hotspots at the outset and then hold those constant for every measurement event (Sherman and Weisburd 1995; Weisburd et al. 2006; Braga and Bond 2008). The corresponding fraction of repeated predictions in such static cases is 1 for each and every day of the experiment.

The fraction of overlapping predictions in Southwest Division are presented in Figure 3 for each combination of treatment and control conditions. On average, 97% of treatment boxes on a focal day are repeated on the immediately subsequent day. Fifteen days later, 25% of the boxes have been replaced with boxes in unique locations, leaving 75% of the boxes in the same places as on the focal day. Thirty days after the initial prediction set, 62% of the initial boxes remain on average. By contrast, only 27% of control boxes on a focal day are repeated on the immediately subsequent control day. The fraction of repeated control boxes declines rapidly, but then rebounds near day 7 suggesting a tendency for analysts to make predictions by day of the week. Thirty days following the initial set of control predictions, only 3% of the same prediction boxes remain on average. Control boxes and treatment boxes follow one another

Table 4. Analyst hotspot selection and hotspot mapping

Analyst	3-Day	7-Day	14-Day	21-Day	28-Day	Remainder
Maidstone	0.105	0.067	0.049	0.033	0.014	0.732
Sevenoaks	0.064	0.073	0.073	0.043	0.022	0.725
Southwest	0.088	0.077	0.054	0.038	0.028	0.716
Foothill	0.073	0.045	0.002	0.004	0.006	0.869
N. Hollywood	0.202	0.154	0.023	0.013	0.008	0.600

NOTE: Fraction of analyst hotspots in agreement with N-day hotspot maps. Those analyst hotspots agreeing with an N-day hotspot map are not included in the fraction for M-day hotspot maps when M > N. Numbers are rounded and may not sum exactly to one.

Table 5. Analyst hotspot selection and near-repeat victimization

Analyst	3-Day	7-Day	14-Day	21-Day	28-Day	Remainder
Maidstone	0.310	0.133	0.100	0.059	0.028	0.369
Sevenoaks	0.239	0.189	0.146	0.065	0.031	0.330
Southwest	0.182	0.120	0.078	0.051	0.040	0.530
Foothill	0.265	0.241	0.041	0.010	0.012	0.430
N. Hollywood	0.307	0.207	0.045	0.019	0.0165	0.405

NOTE: Fraction of analyst hotspots in agreement with N-day hotspot maps or adjacent to boxes selected by an N-day hotspot map. Those analyst hotspots agreeing with or adjacent to an N-day hotspot map are not included in the fraction for M-day hotspot maps when M > N. Numbers are rounded and may not sum exactly to one.

Table 6. Analyst hotspot selection and zero crime rate hotspots

Maidstone	Sevenoaks	Southwest	Foothill	N. Hollywood
0.135	0.117	0.077	0.194	0.139

NOTE: Fraction of analyst hotspots containing no crime throughout the course of the experimental period.

at even lower frequencies. Only about 9% of treatment boxes follow control boxes on an immediately subsequent day, and vice versa. The fraction decays to 4% and 6%, respectively, at 30 days. In general, treatment and control prediction locations are decoupled over time.

3.4 Patrol Activity in Los Angeles

We next consider measures of patrol officer activity in Los Angeles. Like specific field tactics, we expect time on mission to vary across patrol teams in part as a matter of preference, but also that control and treatment conditions should be equivalently exposed to this variation. Figure 4(a) confirms this expectation. Patrol teams spent equivalent amounts of time in control and treatment boxes (Wilcoxon rank-sums test W = 7983, p = 0.8192). This pattern was consistent regardless of whether the patrol team logged large or small numbers of minutes in prediction boxes.

The number of minutes officers spent in a prediction box during a single stop is exponentially distributed up through stops of 120 min in duration ($s(t) = 4360.1e^{-0.049t}$, $R^2 = 0.9943$) (see Figure 4(b)). The mean duration of a directed mission stop was 20.27 min (sdev = 25.11 min). Aggregating all observed mission stops per day, the average time on mission increased in Foothill and Southwest Division over the course of the experiment, and remained stable in North Hollywood (see Figures 5(a)–5(c)). Increases in total time on mission coincided with the concentration of activity in a subset of prediction boxes. Treatment boxes on average received 57 min more exposure to policing per day than control boxes, but the difference is marginally not significant (mean control = 552 min, sdev = 376 min; mean treatment = 617 min, sdev = 430 min; F = 3.288, p = 0.07036).

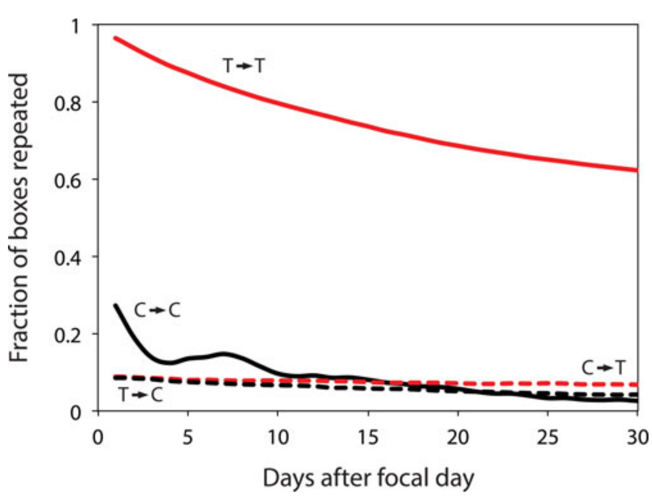

Figure 3. Fraction of prediction boxes repeated on day N following a focal treatment or control day. Comparisons are between predictions delivered for the same shift.

The day-to-day variation in time on mission is high, reflecting the directive to use available time to police prediction boxes. We exploit this variability in patrol time to measure the impact of control and treatment conditions on changes in crime volume in the following section.

3.5 Impact on Crime Rates in Los Angeles

We consider changes in crime volume for all three LAPD divisions as a whole, rather than exclusively within prediction boxes. This is a significant difference with other studies (Sherman and Weisburd 1995; Braga and Bond 2008; Ratcliffe et al. 2011), which rely on random assignment of matched hotspots into treatment and control blocks. A random block design has a disadvantage that one can only detect crime rate changes between matched hotspots, not overall. Because we randomly assign days to treatment and control and, more importantly, allow prediction locations to change twice daily, we are able to measure the overall effects of predictive policing.

Regression models of the form $Y_{ij} = \mu_j + \beta_k X_{ij} + \epsilon_{ij}$ are constructed to assess the relationship between daily crime volume and patrol time on mission. Here Y_{ij} is the daily crime volume on day i in division j, X_{ij} is the police patrol time in minutes on day i in division j across all active prediction boxes, and ϵ_{ij} is the uncorrelated error. The coefficient μ_j is an estimate of the mean crime volume per day in division j in the absence of directed patrol, while β_k is an estimate of the impact of increasing patrol dosage under experimental condition k. The variance in daily crime volume is high for the three focal crime types, ranging from a minimum of 1 to a maximum 18 crimes per day. Nevertheless, crime rates are measurably lower with increasing patrol time in prediction boxes under some conditions. There is a statistically significant negative relationship between crime volume and time on mission for combined control and treatment conditions ($\beta_{\text{combined}} = -7.54 \cdot 10^{-4}$, $p = 0.0208$; Table 6). Treatment conditions alone are also statistically significant ($\beta_{\text{treatment}} = -9.78 \cdot 10^{-4}$, $p = 0.0221$). Control conditions alone display no significant relationship between daily crime volume and patrol time ($\beta_{\text{control}} = -4.66 \cdot 10^{-4}$, $p = 0.364$).

Treatment conditions are expected to yield one less crime per 1000 min of police patrol time in ETAS-predicted locations. The impact translates to 4.3 fewer crimes per week at mean patrol levels, or a reduction of 7.4% on a mean 58.17 crimes per division per week in the absence of patrol. Control conditions would yield less than half the reduction at equivalent patrol levels, assuming statistical significance could be achieved with a larger sample size. The factor 2.1 difference in crime reduction between control and treatment conditions is in agreement with the factor of 1.4 to 2.2 increase in accuracy of ETAS over analyst predictions observed in silent accuracy testing.

In both Kent and Los Angeles, analyst predictive accuracy is a dramatic improvement over a random baseline, with the intelligence-led approach used in Kent outperforming the more exclusive focus on recent crime events followed by LAPD analysts. However, patrol patterns derived from analyst predictions would need to be significantly increased to see a measurable impact on crime. By contrast, ETAS models systematically assess the relative importance of long- and short-term contributions to

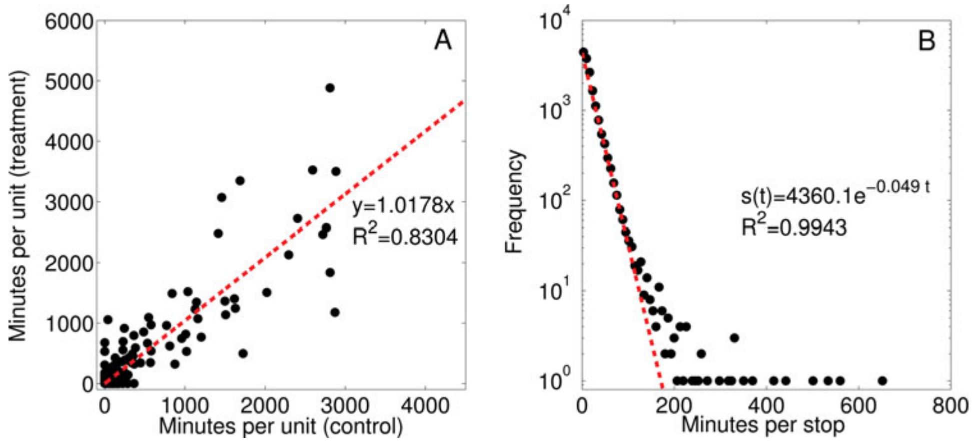

Figure 4. (a) Individual patrol units in LAPD Southwest Division varied in the total amount of time dedicated to prediction missions over the course of the experiment. For example, some patrol units spent nearly 10,000 min in total in prediction boxes, while others spent less than 1000 min in total. The amount of time spent in control and treatment boxes per patrol unit, however, is approximately equivalent. Randomization by day to control or treatment successfully apportioned variation in patrol time engagement among units across experimental conditions. Plotted are cumulative times on treatment and control missions for each patrol unit that was active in Southwest Division between May 20, 2012, and January 10, 2013. The regression line shows that for a patrol unit spending on average 1000 min in control predictions boxes they spent on average 1017 min in treatment prediction boxes. (b) The frequency distribution of minutes per stop in unique 150×150 m prediction boxes. The exponential form of stop durations through 120 min suggests that officer available time is controlled by Poisson arrivals of new calls for service.

dynamic crime risk. Prediction boxes therefore track the probability of where and when crimes are most likely to occur. The distribution of treatment prediction boxes is distinct from both the density of crime and distribution of analyst predictions (Figure 2(c) and 2(f)). Predictive accuracy is approximately doubled and crime reductions are therefore detectable at lower patrol levels.

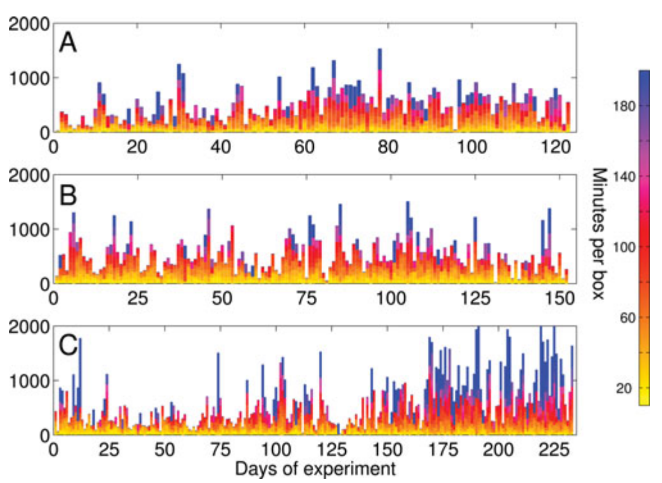

Figure 5. Total minutes on mission per day in (a) Foothill, (b) North Hollywood, and (c) Southwest Divisions. Color represents the distribution of total minutes of patrol in boxes receiving greater (blue) or lower (yellow) amounts of patrol per day. Variation in total minutes on mission reflects day-to-day variation in officer available time, but increases over the course of the experiment in Foothill and Southwest Divisions. The total number of boxes visited per day varies in phase with total minutes on mission. Peak days for total minutes on mission also show concentration of activity in a subset of all boxes. Control and treatment conditions are combined in all three figure panels.

4. DISCUSSION

In this article we introduced an experimental design that allows for the comparison of a short-term crime forecasting algorithm with existing intelligence-led and hotspot policing practices. Hotspot locations change daily based upon updated forecasts and police patrols respond dynamically. Our results serve to validate the premise of predictive policing that crime rates can be reduced under fixed policing resources by increasing the crime rate in hotspots flagged for directed patrol. This is accomplished by using an algorithmic approach to hotspot selection, systematically estimating the relative contributions to risk of short-term and long-term crime patterns.

While the focus in this study was to compare algorithmic approaches to those employed by analysts, we note that the methods also outperform standard hotspot mapping techniques in retrospective forecasts. The accuracy of statistical models of crime hotspots has been compared to that of hotspot maps in Wang and Brown (2012), where generalized additive models are constructed for robbery prediction, and Mohler (2014), where chronic homicide hotspot maps are compared to marked point process models in terms of accuracy. In both studies the statistical models are shown to significantly outperform standard hotspot mapping approaches, where in Mohler (2014) and Equation (1) chronic hotspot mapping is a submodel. Statistical estimation allows one to rigorously estimate the contributions of both chronic and short-term risk to improve model accuracy and at the same time make the analyst's job easier through automation of the model. The partitioning of the probabilistic rate of crime $\lambda_n(t)$ into stationary and dynamic components captures implicitly the roles that environmental heterogeneity and contagion-like processes play in generating crime risk (Johnson 2008). In this sense, our modeling approach parallels methods that focus on contagion (Johnson et al. 2009) and risk facilities (Eck, Clarke, and Guerette 2007; Bowers 2014). However, it is critical to note that these parallels are only *implicit* mechanisms in the ETAS model. Future research should evaluate whether explicit treatment of environmental heterogeneity and contagion within an algorithmic framework would improve performance.

The present study suggests that ETAS crime forecasting outperforms crime forecasting based on intelligence-led and hotspot mapping models. The results find a balance between ecological and external validity. Unlike controlled tests that pit different algorithms against historical data in sterile conditions (Chainey, Tompson, and Uhlig 2008; Mohler et al. 2011), the present experiment was conducted head-to-head with practitioners under all the pressures that confront crime analysts and police on a day-to-day basis. Without such ecological validity, practitioners would never believe that an algorithmic approach could match, let alone outperform the human expert using existing best practice. Such appears to be the case here. Yet this ecological validity in the specific settings of the experimental trials could compromise the ability to generalize the results. Much depends on how representative the practices of the Los Angeles and Kent Police Department are in the broader policing world. While there is considerable tactical and strategic variation in policing practice across departments, the use of crime mapping, hotspot maps, and simple criminal intelligence to identify chronic offenders is as close to universal as anything. In a survey of U.S. police departments, for example, 100% of departments serving populations of 500,000 people or more relied regularly on these analytical methods (Friedmann, Rosenfeld, and Borissova 2010; Reaves 2010). The proportion of police departments falls to 56% for those serving populations of 50,000 to 100,000. Police departments in England and Wales ranked crime mapping in support of intelligence-led policing at the top of the analytical tasks they perform as a proportion of time (Weir and Bangs 2007). The comparison of algorithmic predictions to best practices in intelligence-led and hotspot mapping by two prominent departments therefore would seem to have reasonable external validity especially as department size increases.

More challenging is an assessment of whether similar effect sizes can be expected in other settings. Comparisons between regions based on existing studies are problematic (Bowers, Johnson, and Pease 2004). Tables 1 and 2 show that area standardized accuracies (PAI) across regions vary widely. In the present study area-standardized accuracies range from 3.5 to 85.2 for ETAS and 1.2 to 50.8 for analysts in Southwest and Maidstone, respectively. Variation in area-standardized accuracies is driven by differences in the physical sizes of the regions in question (Tompson and Townsley 2010), the mixtures of crimes included in predictions (Chainey, Tompson, and Uhlig 2008), and the time scales over which predictions are rendered (Tompson and Townsley 2010). Comparison of the relative performance of different forecasting methods within regions helps solve some of these problems. Since different forecasting methods tested within a region are subject to the same constraints of physical area, mix of crimes, and forecasting time scales, the gap between their performance is less subject to such confounding factors. This is the approach followed by Chainey, Tompson, and Uhlig (2008) where kernel density estimation yields an average boost over other hotspotting methods of 1.7 to 1.8. As shown here and in Mohler et al. (2011), ETAS provides a boost over

Table 7. Parameter estimates for multiple regression

	μ_{Foothill}	$\mu_{\text{N. Hollywood}}$	$\mu_{\text{Southwest}}$	β_k	SE
Combined conditions	6.59***	9.36***	8.98***	$-7.54 \cdot 10^{-4}$*	0.000326182
Treatment (ETAS)	6.79***	9.35***	9.12***	$-9.78 \cdot 10^{-4}$*	0.000427327
Control (analyst)	6.60***	9.2***	8.79***	$-4.66 \cdot 10^{-4}$	0.000513346

NOTE: μ_j is the estimated mean crime rate per day in each division, and β_k is the estimated average change in crime volume per minute of patrol time under different experimental conditions k. Standard errors are for estimated β_k. *$p < 0.05$ and ***$p < 0.001$ levels.

kernel density estimation, providing a basis for positing an expected magnitude of results elsewhere. Ultimately, whether this hypothesis proves true is an empirical question that will require further experimental and comparative work.

Beyond the question of broader applicability, we might also ask whether the observed impact on crime is significant as a practical matter. At mean patrol levels (31 min per box per day), the ETAS algorithm corresponds to 4.3 fewer crimes per week per division. While a crime reduction of 7.4% may appear small, the potential savings to society are quite large when estimated using the methodology of McCollister, French, and Fang (2010). In Table 8 we display the societal costs per crime calculated in McCollister, French, and Fang (2010), broken down by costs to the victim, police and court system, and offender (2013 dollars). We then use the relative frequencies of GTA, burglary theft from motor vehicle, and burglary (0.292, 0.584, 0.184) along with the average crime rate per LAPD division at zero patrol levels to estimate the weekly cost of crime per division without predictive policing (Table 9). We then estimate the savings per week achieved by a 4.3 crime reduction per division on average and extrapolate the savings across LAPD. We project a $17,258,801 annual savings to LAPD if ETAS were to be used for 31 min per day in each hotspot compared to no patrol. Patrols allocated under analyst conditions are projected to achieve less than half the savings ($8,223,519).

From an ethnographic perspective, we believe predictive policing operates at a local level through short-term disruption of criminal opportunities. A representative scenario arises when an officer shows up at a location designated as high risk. An offender who lives or works in that area (Brantingham and Brantingham 1991) sees the officer and decides to lay low or even run (Goffman 2014). In that time they are laying low they are in no position to commit a crime. If the offender comes out a few hours later and again sees the officer in the same or a nearby hotspot, the deterrence effect may last well beyond those particular policing events. Short-term crime prevention of this type is different from long-term crime prevention meth-

Table 8. Cost per crime

Offense type	Victim cost	Police & court cost	Offender cost	Total per crime
GTA	$6615	$4184	$598	$11,398
Burglary	$1474	$4465	$737	$6676
BTFV	$519	$3115	$176	$3812
Total Costs	$8608	$11,765	$1512	$21,886

NOTE: Cost of crime for the crime types targeted in the LAPD experiment calculated from McCollister, French, and Fang (2010) converted into 2013 dollars.

Table 9. Cost savings per LAPD division

Offense type	Divisional cost per week w/o ETAS	Divisional savings per week w/ETAS	Annual savings per division	Annual savings LAPD (projected)
GTA	$70,971	$5246	$272,808	$5,728,959
Burglary	$47,845	$3537	$183,913	$3,862,182
Theft	$94,988	$7022	$365,127	$7,667,660
Total Costs	$213,805	$15,805	$821,848	$17,258,801

NOTE: Savings per LAPD division calculated using the methodology of McCollister, French, and Fang (2010) and a mean reduction of 4.3 crimes per division per week under ETAS patrol.

ods aimed at fixing the root causes of crime such as building positive police–community relations, combating chronic drug dependencies, or changing the built environment to reduce risk. We are not arguing that predictive policing is a singular approach, nor is it a replacement for other proven crime prevention strategies. Rather, accurate predictions about where and when crime is most likely to occur simply enhance the ability of police officers (compared to social workers, counselors, teachers, etc.) to achieve short-term crime prevention more effectively.

In future work, greater crime reductions may be achieved by improving predictive algorithms. For this purpose, maximum likelihood estimation may not be the best method for optimization. Nonsmooth loss functions such as precision are likely more relevant for predictive policing, where the goal is to maximize the percentage of crime in *k* flagged hotspots. Learning to rank methods from information retrieval (Burges 2010) can potentially be used for this purpose to prevent risk estimates in the top-ranked hotspots from being biased by the influence of medium and low-risk hotspots, as is the case with maximum likelihood.

The experimental protocol used here is similar to a repeated measures crossover design wherein the same experimental subjects are exposed to different treatment conditions. Here each day is randomly assigned to control and treatment conditions and thus control and treatment events can follow one another in different sequences. We have assumed independence of days. Nevertheless, it is possible that patrol in an ETAS hotspot on one day may influence crime in an analyst hotspot on the next day (or vice versa). There is some evidence that patrol effects last on a shorter time scale ($\sim$2 hr) (Koper 1995) and that displacement or diffusion of benefits in time may not extend out to a day or longer (or the effect may be small). The random allocation of experimental conditions by day helps to alleviate some of this concern. In general, control days are equally likely to be followed by another control day or a treatment day (and vice versa). Thus, treatment and control days experience equal exposure to any possible contamination effects. Future experiments might deploy buffer days or randomization by weeks to further separate treatment and control conditions in time. Longer experimental periods would be needed to achieve similar significance to this study, however. Alternatively, statistical corrections might be developed if contamination across conditions was clearly indicated.

More research is also needed to better understand the relationship between police patrol and crime. Whereas officer call logs were used in this study, in-car GPS can provide a greater level of precision and also provide information on officer activity when not on a predictive policing mission. Furthermore, given the deterrent effect observed in our field experiments, patrol times and locations may be important variables to include in predictive policing models to close the feedback loop between officers, criminals, and the algorithms linking them together.

[Received June 2014. Revised June 2015.]

REFERENCES

Bowers, K. (2014), "Risky Facilities: Crime Radiators or Crime Absorbers? A Comparison of Internal and External Levels of Theft," *Journal of Quantitative Criminology*, 30, 389–414. [1408]

Bowers, K. J., Johnson, S. D., and Pease, K. (2004), "Prospective Hot-Spotting the Future of Crime Mapping?" *British Journal of Criminology*, 44, 641–658. [1400,1402,1409]

Braga, A. A. (2005), "Hot Spots Policing and Crime Prevention: A Systematic Review of Randomized Controlled Trials," *Journal of Experimental Criminology*, 1, 317–342. [1399]

Braga, A. A., and Bond, B. J. (2008), "Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial," *Criminology*, 46, 577–607. [1400,1401,1406,1407]

Brantingham, P. J., and Brantingham, P. L. (1981), *Environmental Criminology*, Beverly Hills, CA: Sage. [1399]

Brantingham, P. L., and Brantingham, P. J. (1991), "Notes on the Geometry of Crime," in *Environmental Criminology*, Thousand Oaks, CA: Sage, pp. 27–54. [1409]

Burges, C. J. C. (2010), "From Ranknet to Lambdarank to Lambdamart: An Overview," *Learning*, 11, 23–581. [1410]

Chainey, S., and Ratcliffe, J. (2005), *GIS and Crime Mapping*, Chichester, UK: Wiley. [1403]

Chainey, S., Tompson, L., and Uhlig, S. (2008), "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime," *Security Journal*, 21, 4–28. [1400,1404,1409]

Eck, J. E., Clarke, R. V., and Guerette, R. T. (2007), "Risky Facilities: Crime Concentration in Homogeneous Sets of Establishments and Facilities," *Crime Prevention Studies*, 21, 225. [1408]

Farrell, G., and Pease, K. (1993), *Once Bitten, Twice Bitten: Repeat Victimisation and Its Implications for Crime Prevention*, London: Police Research Group, Crime Prevention Unit. [1406]

Farrington, D. P., and Welsh, B. C. (2005), "Randomized Experiments in Criminology: What Have We Learned in the Last Two Decades?" *Journal of Experimental Criminology*, 1, 9–38. [1399]

Fielding, M., and Jones, V. (2012), "'Disrupting the Optimal Forager': Predictive Risk Mapping and Domestic Burglary Reduction in Trafford, Greater Manchester," *International Journal of Police Science & Management*, 14, 30–41. [1400]

Friedmann, R. R., Rosenfeld, R., and Borissova, N. (2010), "Improving Crime Data Project," Technical Report, U.S. Department of Justice. [1403,1409]

Goffman, A. (2014), *On the Run: Fugitive Life in an American City*, Chicago, IL: University of Chicago Press. [1409]

Guerette, R. T., and Bowers, K. J. (2009), "Assessing the Extent of Crime Displacement and Diffusion of Benefits: A Review of Situational Crime Prevention Evaluations," *Criminology*, 47, 1331–1368. [1399]

Harrison, G. W., and List, J. A. (2004), "Field Experiments," *Journal of Economic Literature*, 42, 1009–1055. [1400]

Heckman, J. J. (2008), "Econometric Causality," *International Statistical Review*, 76, 1–27. [1400]

Johnson, S. (2008), "Repeat Burglary Victimisation: A Tale of Two Theories," *Journal of Experimental Criminology*, 4, 215–240. [1408]

Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., and Townsley, M. (2007), "Space–Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization," *Journal of Quantitative Criminology*, 23, 201–219. [1399]

Johnson, S. D., Birks, D. J., McLaughlin, L., Bowers, K. J., and Pease, K. (2007), *Prospective Crime Mapping in Operational Context: Final Report*, London: Great Britain Home Office Research Development and Statistics Directorate. [1400]

Johnson, S. D., Bowers, K. J., Birks, D. J., and Pease, K. (2009), "Predictive Mapping of Crime by Promap: Accuracy, Units of Analysis, and the Environmental Backcloth," in *Putting Crime in Its Place*, New York: Springer, pp. 171–198. [1400,1408]

Keizer, K. (2008), "Siegwart Lindenberg, and Linda Steg. The Spreading of Disorder," *Science*, 322, 1681–1685. [1399]

Kelling, G. L., Pate, T., Dieckman, D., and Brown, C. E. (1974), *The Kansas City Preventive Patrol Experiment*, Washington, DC: Police Foundation. [1399]

Kennedy, D. M. (1996), "Pulling Levers: Chronic Offenders, High-Crime Settings, and a Theory of Prevention," *Valparaiso University Law Review*, 31, 449. [1404]

Kennedy, L. W., Caplan, J. M., and 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, 339–362. [1400]

Kirk, R. E. (1982), *Experimental Design*, Belmont, CA: Brooks/Cole. [1400,1401]

Koper, C. (1995), "Just Enough Police Presence: Reducing Crime and Disorderly Behavior by Optimizing Patrol Time in Crime Hot Spots," *Justice Quarterly*, 12, 649–672. [1401,1410]

Lewis, E., and Mohler, G. (2011), "A Nonparametric EM Algorithm for Multiscale Hawkes Processes," in *Proceedings of the 2011 Joint Statistical Meetings*, pp. 1–16. [1402]

Liu, H., and Brown, D. E. (2003), "Criminal Incident Prediction Using a Point-Pattern-Based Density Model," *International Journal of Forecasting*, 19, 603–622. [1400]

Marsan, D., and Lengline, O. (2008), "Extending Earthquakes' Reach Through Cascading," *Science*, 319, 1076–1079. [1400,1401]

Matsueda, R. L., Kreager, D. A., and Huizinga, D. (2006), "Deterring Delinquents: A Rational Choice Model of Theft and Violence," *American Sociological Review*, 71, 95–122. [1399]

McCollister, K. E., French, M. T., and Fang, H. (2010), "The Cost of Crime to Society: New Crime-Specific Estimates for Policy and Program Evaluation," *Drug and Alcohol Dependence*, 108, 98–109. [1409]

Mohler, G. (2014), "Marked Point Process Hotspot Maps for Homicide and Gun Crime Prediction in Chicago," *International Journal of Forecasting*, 30, 491–497. [1400,1401,1402,1404,1408]

Mohler, G., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011), "Self-Exciting Point Process Modeling of Crime," *Journal of the American Statistical Association*, 106, 100–108. [1400,1401,1402,1404,1409]

Mosher, C. J., Miethe, T. D., and Hart, T. C. (2010), *The Mismeasure of Crime*, Thousand OaksCA: Sage. [1404]

National Centre for Policing Excellence (2005), *Guidance on the National Intelligence Model*, London: National Centre for Policing Excellence, United Kingdom. [1403,1404]

O'Shea, T. C., Nicholls, J. K., Archer, J., Hughes, E., and Tatum, J. (2003), *Crime Analysis in America: Findings and Recommendations*, Washington, DC: U.S. Department of Justice, Office of Community Oriented Policing Services. [1403]

Ratcliffe, J. H. (2012), *Intelligence-Led Policing*, New York: Routledge. [1403]

Ratcliffe, J. H., Taniguchi, T., Groff, E. R., and Wood, J. D. (2011), "The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots," *Criminology*, 49, 795–831. [1399,1400,1401,1407]

Reaves, B. A. (2010), *Local Police Departments, 2007*, Bureau of Justice Statistics Report NCJ 231174, Washington, DC: Bureau of Justice Statistics. [1403,1409]

Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997), "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy," *Science*, 277, 918–924. [1399]

Sherman, L. W., Gartin, P. R., and Buerger, M. E. (1989), "Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place," *Criminology*, 27, 27–56. [1399]

Sherman, L. W., and Weisburd, D. (1995), "General Deterrent Effects of Police Patrol in Crime? Hot Spots?: A Randomized, Controlled Trial," *Justice Quarterly*, 12, 625–648. [1399,1401,1403,1406,1407]

Short, M. B., Brantingham, P. J., Bertozzi, A. L., and Tita, G. E. (2010), "Dissipation and Displacement of Hotspots in Reaction-Diffusion Models of Crime," *Proceedings of the National Academy of Sciences*, 107, 3961–3965. [1399,1401]

Telep, C. W., Mitchell, R. J., and Weisburd, D. (2014), "How Much Time Should the Police Spend at Crime Hot Spots? Answers From a Police Agency Directed Randomized Field Trial in Sacramento, California," *Justice Quarterly*, 31, 905–933. [1401]

Tompson, L., and Townsley, M. (2010), "(Looking) Back to the Future: Using Space-time Patterns to Better Predict the Location of Street Crime," *International Journal of Police Science & Management*, 12, 23–40. [1409]

Townsley, M., Homel, R., and Chaseling, J. (2003), "Infectious Burglaries: A Test of the Near Repeat Hypothesis," *British Journal of Criminology*, 43, 615–633. [1406]

Veen, A., and Schoenberg, F. P. (2008), "Estimation of Space-Time Branching Process Models in Seismology Using an EM-Type Algorithm," *Journal of the American Statistical Association*, 103, 614–624. [1402]

Walsh, W. F. (2001), "Compstat: An Analysis of an Emerging Police Managerial Paradigm," *Policing: An International Journal of Police Strategies & Management*, 24, 347–362. [1403]

Wang, X., and Brown, D. E. (2012), "The Spatio-Temporal Modeling for Criminal Incidents," *Security Informatics*, 1, 1–17. [1400,1404,1408]

Weir, R., and Bangs, M. (2007), *The Use of Geographic Information Systems by Crime Analysts in England and Wales*, London: Great Britain Home Office Research Development and Statistics Directorate. [1403,1409]

Weisburd, D. (2008), "Place-Based Policing," *Ideas in American Policing*, 9, 1–15. [1399]

Weisburd, D., and Eck, J. E. (2004), "What Can Police Do to Reduce Crime, Disorder, and Fear?" *The Annals of the American Academy of Political and Social Science*, 593, 42–65. [1399]

Weisburd, D., Wyckoff, L. A., Ready, J., Eck, J. E., Hinkle, J. C., and Gajewski, F. (2006), "Does Crime Just Move Around the Corner? A Controlled Study of Spatial Displacement and Diffusion of Crime Control Benefits," *Criminology*, 44, 549–592. [1399,1406]