Redrawing Hot Spots of Crime in Dallas, Texas*

Andrew P. Wheeler^{1,**}, Sydney Reuter²

*Data and code to replicate the findings can be obtained from
<https://www.dropbox.com/sh/kcask6pinaaaz4v/AAC4CXk6NzUweyld2n4OznzWa?dl=0>

**Corresponding author, email: apwheele@gmail.com

1. HMS

2. School of Economic, Political, and Policy Sciences – Program in Criminology & Criminal Justice

Andrew P. Wheeler, PhD, is a Data Scientist at HMS. This work does not represent the views of HMS.

Sydney Reuter is a Masters student in the Criminology program at UT Dallas.

Title: Redrawing hot spots of crime in Dallas, Texas

Abstract: In this work we evaluate the predictive capability of identifying long term, micro place hot spots in Dallas, Texas. We create hot spots using a clustering algorithm, using law enforcement cost of crime estimates as weights. Relative to the much larger current hot spot areas defined by the Dallas Police Department, our identified hot spots are much smaller (under 3 square miles), and capture crime cost at a higher density. We also show that the clustering algorithm captures a wide array of hot spot types; some one or two addresses, some street segments, and others an agglomeration of larger areas. This suggests identifying hot spots based on a specific unit of aggregation (e.g. addresses, street segments), may be less efficient than using a clustering technique in practice.

Keywords: hot-spots, clustering, prediction, cost-benefit-analysis

Code and data to reproduce the analysis can be downloaded from

<https://www.dropbox.com/sh/kcask6pinaaaz4v/AAC4CXk6NzUweyld2n4OznzWa?dl=0>

Introduction

While targeting police resources at micro place hot spots has been one of the most successful policing interventions to reduce crime (Braga, Turchan, Papachristos, & Hureau, 2019), it is still an open question as to how to construct those hot spots. Using open crime data from Dallas, Texas, this article provides an example of constructing long term micro place hot spots. Using historical data to identify hot spots of crime, we show how our identified clusters capture a higher density of crime costs than current Dallas hot spot areas on future crimes.

While there are a plethora of prior articles examining the predictive ability of different hot spot methods (Chainey, Thompson, & Uhlig, 2008; Drawve, 2016; Levine, 2008; Van Patten, McKeldin-Conor, & Cox, 2009), there are three novel contributions of this work. First, we illustrate the use of a hierarchical clustering technique, DBSCAN (Campello, 2013), to formulate different contiguous hot spot areas. A point of contention in prior work is the correct spatial unit of analysis to conduct analysis on. For example, several scholars often suggest street segments (Rosser et al., 2017; Weisburd et al., 2004), some suggest street segments and intersections (Braga, Papachristos, & Hureau, 2010; Wheeler, Worden, & McLean, 2016), and others have suggested targeting specific addresses (Eck, Clarke, & Guerette, 2007; Lee & Eck, 2019; Sherman, Gartin, & Buerger, 1989). Using a hierarchical clustering technique allows one to avoid needing to specify such a spatial unit of analysis up front, and can be used on address based crime data to identify a reasonable resolution for a particular hot spot. In the micro place hot spots we identify, we subsequently show some are one or two addresses clustered together, others are a street segment, and others are an agglomeration of several nearby street segments.

The second novel contribution of this work is to construct hot spots using weighted cost of crime scores, which is related to prior research creating hot spots using crime harm scores

(Ignatans & Pease, 2016; Ratcliffe 2015; Sherman, Neyroud, & Neyroud, 2016; Macbeth & Ariel, 2017). Crime harm has been characterized in prior research by either asking survey respondents to rank different crimes (Wolfgang, 1985), or by translating sentencing decisions to create harm weights (Ratcliffe, 2015), with the ultimate goal that policing resources are better allocated relative to the harm a particular crime impacts on the community, as opposed to counting all crime equal (Sherman & Cambridge University Associates, 2020).

In place of crime harm scores however, here we use average Uniform Crime Report (UCR) cost of crime estimates (per law enforcement) for Texas as weights in the DBSCAN algorithm (Hunt, Saunders, & Kilmer, 2019). This provides more interpretable hot spot summaries in terms of direct cost of crime estimates relevant to police departments. These cost of crime estimates can be used to provide more actionable information in terms of cost-benefit analysis for police departments when planning a hot spots policing strategy.

The third novel contribution of the study is to evaluate the predictive accuracy of our identified crime weighted hot spots using historical data to predict future crimes. While prior work has mapped harm spots (Curtis-Ham & Walton, 2017; Fenimore, 2019; Norton et al., 2018; Weinborn et al., 2018), the majority of that work has simply focused on crime concentration, and has not assessed the predictive accuracy of a technique to create such harmspots (for an exception, see Macbeth & Ariel, 2017). Additionally, most of that work has mapped harm spots at a specific pre-chosen unit of analysis (Curtis-Ham & Walton, 2017; Mitchell, 2019; Norton et al., 2018; Weinborn et al., 2018, for an exception see Fenimore, 2019 who used weighted kernel density estimates). We show here in our constructed hot spot areas there is no spatial unit of analysis that would consistently cover the same areas we identified.

The study area of Dallas was chosen for one main reason. They have historically identified long term hot spot areas, Target Area Action Grids (TAAG) (Ferguson, 2017; Worrall & Wheeler, 2019). These hot spot areas on their face do not conform with contemporary advice on targeting micro place hot spots of crime. TAAG areas cover 64 square miles of Dallas, around 20% of the city. We show that TAAG areas capture around 60% of the total crime cost in the city. Weisburd's law of crime concentration states that approximately 5% of the places within a city typically capture around 50% of the total amount of crime (Weisburd, 2015). Thus it suggests one can identify a much smaller subset of the city and capture crime at a much higher density.

A common statistic used to evaluate crime forecasts is the Predictive Accuracy Index (PAI) (Chainey et al., 2008). The PAI is simply the ratio of the proportion of crime captured for the numerator and the percent of the area covered by the prediction in the denominator. A simple extension this statistic is to consider not just the proportion of total crimes captured, but the total cost of crime captured. We show that current TAAG areas in Dallas have cost-of-crime PAI's of around 3 (60% crime cost/20% of area of city), while Weisburd's law of crime concentration suggests a reasonable baseline for PAI statistics should be 10 (50% of crime cost/5% of the area). Our identified hot spot areas are much smaller, cover less than 1% of the area of the city (around 2.5 square miles), capture 12% of the total cost of crime in the city, and subsequently have a combined PAI of 16.

Literature Review

Creating Hot Spots of Crime

Work that identified that crime clusters at small locations, such as specific addresses (Sherman, Gartin, & Buerger, 1989), or street segments (Weisburd et al., 2004) was likely the

innovation that spurred subsequent work on hot spots policing (Braga et al., 2019; Lum, Koper, & Telep, 2011). This innovation was particularly important given the lack of prior successes of police interventions to reduce crime (Bayley, 1994). In retrospect, those poor results are likely attributable to police targeting of much larger areas (Larson, 1975), which has been consistently shown to be less successful than targeting micro place hot spots (Lum, Koper, & Telep, 2011).

However, despite the popularity of hot spots policing, there is not a single, unanimous definition of what constitutes a hot spot of crime (Taylor, 2015). This is because a hot spot is ultimately a data based definition, not a physical entity that one can go and show in the physical world. Thus while it is easy to show that a small number of micro locations have a relatively large number of crime counts (Weisburd, 2015), it is much harder to delineate the boundaries of what is inside or what is outside of a hot spot (Taylor, 2015). Subsequently there have been a myriad of different ways using data to define hot spots.

Sherman, Gartin, & Buerger (1989) conducted a study on hot spots in Minneapolis which confirmed that relatively few areas produce the most calls to police. They found that half of all calls to police were in only 3% of the places (addresses or intersections) within the city. Subsequent work by David Weisburd has stated a consistent law of crime concentration, that around 5% of the places in a city (where places are often defined as street segments) contain around 50% of the crime (Weisburd, 2015). These approaches to identifying hot spots rely on one to pre-specify a particular unit of analysis (e.g. address, street segment, census block, patrol area, etc.), as well as a particular ranked threshold (e.g. the top 5% of areas, or the top 10 areas) to set a cut-off for what constitutes a hot spot.

Chainey et al. (2008) identified a number of different mapping techniques that can be used to identify hot spots of crime that don't *per se* rely on a pre-defined unit of analysis or a

simple threshold (e.g. top 5% of the area). Examples include point mapping, clustering census areas, or kernel density estimation. The depiction of the hot spots on a map may vary depending of the level of concentration for the analysis. For example, if the focus is on streets, lines may be used, while if it was on areas/neighborhoods, shaded areas or gradients may be more appropriate (Eck et al., 2005). While these techniques potentially avoid relying on some arbitrary decisions (such as a spatial unit of analysis), they still rely on other arbitrary decisions, such as where to set the cut-off for a kernel density threshold to set the area for a hot spot. Frequently individuals use rules such as a z-score above 3 to make such decisions for continuous predictions (Eck et al., 2005; Chainey et al., 2008; Drawve, 2016).

According to Haberman (2017), there are several different ways to identify hot spots. Before beginning, the size of the hot spot areas needs to be established. This can range from a large police beat, to a smaller grid area, and even to an even smaller microplace consisting of an intersection or street segment. It is important to then determine the criteria that may qualify an area as a hot spot. A common method includes the establishment of a minimum number of crimes that must have occurred in that area. Separately, selecting the locations with the highest number of crimes after rank-ordering the areas can be used to establish hot spots. Finally, spatial statistics may be used to identify clusters of crime, therefore revealing hot spots. Utilizing one of these methods, most researchers have identified hot spots of all crime within an urban area, while others have focused on a specific type of crime, such as robbery or motor vehicle thefts (Haberman, 2017).

Given that how to draw hotspots is arbitrary, there has been extensive work on analyzing the accuracy of different hot spot mapping techniques (Chainey et al., 2008; Drawve, 2016;

Flaxman et al., 2019; Lee et al., 2019; Mohler & Porter, 2018). Here we focus on one of the more common metrics, the *Predictive Accuracy Index* (PAI) (Kounadi et al., 2020).

Measuring Hot Spot Accuracy using the Predictive Accuracy Index

The Prediction Accuracy Index (PAI), is a technique used to calculating the overall accuracy of a hot spot, but take into account the size of the area predicted (Chainey et al., 2008; Drawve, 2016). For example, if one simply used the number of crimes captured in a predicted area, or similarly the *hit rate* of whether any crime occurred in a predicted area (for small temporal intervals), it would be unfair to compare a prediction that encompassed a much larger area than a smaller predicted hot spot. Using the hit rate, you would gain the best prediction by predicting the entire study area, which is ultimately not helpful in identifying small hot spots of crime.

The PAI metric however divides the total crimes captured in an identified hot spot with the size of the area being considered. The PAI formula can be written as:

$$\frac{n/N}{a/A} = \frac{\% \text{ Crime Captured}}{\% \text{ Area of City}} = \text{PAI}$$

Where n is the number of crimes captured in the identified hot spot, and N is the total number of crimes in the entire study area, and this is the numerator for the PAI metric. The denominator is defined by a , the area of the hot spot, divided by A , the area of the entire city. As such, the metric reduces to the proportion of crimes captured in a hot spot for the numerator, and the proportion of the area considered in the denominator. For a simple example, Weisburd's law of crime concentration would suggest that 5% of the area of the city approximately captures 50% of the crime, which results in a PAI statistic of $0.5/0.05 = 10$.

It is important when evaluating this metric to use prospectively identified hot spots, and then use future crimes captured in those previously defined areas, e.g. create hot spots using 2016 data, and see how many crimes they capture in 2017 data. If one uses the same dataset to evaluate both, the accuracy of the method used to generate the hot spots will be greatly inflated (Berk, 2008).

Given we are interested in capturing weighted crime hot spots, we slightly adapt the PAI metric to take those weights into account. If one defines the weighted crime cost (or harm) captured in a hot spot and the weighted crime cost (or harm) in an entire city as below, where w indicates the weight for a particular crime:

$$\begin{aligned}\sum w \cdot n &= v, \text{ Weighted crime captured in hot spot} \\ \sum w \cdot N &= V, \text{ Weighted crime in entire study area}\end{aligned}$$

We can then simply replace the numerator in the prior PAI formula with v/V , and have a value weighted PAI_v statistic.

$$\frac{v/V}{a/A} = \frac{\% \text{ Crime Harm Captured}}{\% \text{ Area of City}} = PAI_v$$

This statistic is agnostic as to what constitutes the weights, which segues into our next section, how prior research has quantified crime cost (or harm) weights to capture the notion that some crimes (e.g. violent interpersonal offenses) should take priority over lower level offenses (e.g. petty theft) when identifying areas for police departments to target.

Calculating Crime Harm Scores

Recognizing that all types of crime count the same amount towards the crime rate, although they differ in severity, Wolfgang (1985) conducted a survey to determine how the public perceived the severity of crimes. Each respondent was asked to compare and rate a

number of crimes based on their description and details of the incident. These ratings were averaged to determine an overall score for each crime. Patterns could be seen in the scores from the survey, such as crimes involving multiple victims typically receiving a higher severity score than the same crimes with a fewer number of victims. Additionally, factors such as who the victim was, the amount of money involved in the crime, or the original intent of the individual, are all additional factors that were important to respondents when assigning severity scores.

Many researchers have begun developing additional harm indexes to more appropriately measure crime than by a count or rate within an area. The majority of these harm indexes are based on sentencing data, as opposed to using public opinion surveys. Although the methods to create such harm indexes have varied slightly over different locations.

Ratcliffe (2015) attempted to create a realistic measure of harm experienced by the community due to crime. Focusing in Philadelphia, each offense was assigned a score based on general guidelines to assist judges in assigning an appropriate penalty for guilty individuals. From here, the author believed the index should also include harms suffered by the community that were not included in the harm index since it focuses mainly on violent crime. This 'social harm perspective' extends to include poverty/cash-loss, fraud, emotional harm, psychological harm, or sexual harm that may not be captured by serious crime numbers. The addition of other crime types, traffic accidents, and investigative activity/stops is more reflective of community harm and police responsibilities.

Sherman, Neyroud, & Neyroud (2016) created the Cambridge Crime Harm Index, adopted for the UK from the previously created Crime Harm Index, where they rated the harm of each crime reported to the police. This measure excluded any proactive crime detections from the police department, as they could not reliably measure harm experienced by the population.

Each crime was multiplied by the number of days in prison an offender would receive upon conviction of that crime.

The New Zealand Crime Harm Index is another example of measuring the harm of crime, in an effort to reallocate police resources more efficiently. Curtis-Ham & Walton (2018) developed this measure as an alternative to the traditional Crime Harm Index, to be more specific with crimes in New Zealand. Although it takes a similar approach, the New Zealand Crime Harm Index uses the actual time served by the offender, to maintain proportionality between short and long sentences due to early releases and parole, account for Home Detention allotment, as well as account for the proportion of the sentence served.

The Western Australia Crime Harm Index was developed by House & Neyroud (2018) to analyze harm levels among common offenses for first-time offenders. It was calculated by taking the median number of days sentenced for each first time offender in a sample of most common offense categories. Conditional community sentences and monetary penalties were additionally accounted for in terms of the equivalent number of prison days. The number of offenses for each category was multiplied by the median number of prison days sentenced per category. This created the Western Australia Crime Harm Index, which was found to have the potential to improve policing, as well as community safety in Western Australia.

Mapping crime harm spots can take many different forms. Weinborn, Ariel, Sherman, & O'Dwyer (2017) identified the top 5% of street segments (two standard deviations away from the mean) that accounted for the most crime in the city to be hotspots, and constructed weights based on a crime harm index centered around sentencing data. Curtis-Ham & Walton (2017) used census units to determine which areas of New Zealand had the highest crime harm, specifically comparing the Crime Harm Index with the New Zealand Priority Locations Index. Norton, Ariel,

Weinborn, & O'Dwyer (2018) measured harm within street segments using the Crime Harm Index based on sentencing guidelines, and included results within three standard deviations of the mean. They identified the four most harmful offense categories (violence against persons, sexual offense, robbery and theft) to make up 80% of the harm within the 99 harm spots.

Mitchell (2019) created the California Crime Harm Index using a measure of potential maximum prison days to evaluate the Sacramento Hot Spot Experiment in comparison with crime counts for the area. Fenimore (2019) used weighted kernel density estimates to show that while crime harm tended to be as concentrated as crime counts, crime clusters were in different areas than clusters with crime counts. This was due to violent crimes contributing more to the weights, which were based on sentencing guidelines or the prior severity scores based on the Wolfgang survey.

As opposed to using crime harm estimated via sentencing decisions, we take a different tact in this analysis; we use estimates of costs of crime relevant directly to a police department. Hunt, Saunders, and Kilmer (2019) chose to calculate crime specific variable cost estimates (for police departments) to aid in cost-benefit analysis. Using the labor costs of crime to construct hot spots, instead of focusing on sentencing guidelines and punishment, can directly relate to the benefits of reducing crime in that area that are directly realized by police.

To briefly describe Hunt et al.'s (2019) methodology in more detail, it specifically focuses on the labor costs of *responding* to crime. Criminologists often distinguish between reactive and proactive policing (Nagin et al., 2015), and Hunt's estimates are relevant for the reactive category. The reactive police response to crime can be broken down in-between various categories; Hunt et al. (2019) list different categories administrative, arrest, crime scene, court time, investigative, and en-route/waiting. They generate state level estimates for costs of

responding to crime based on various data sources, including Bureau of Justice Statistics resources for justice expenditures and time spent responding to different crimes (broken down by differing roles, such as patrol officer vs detective), and state crime incident totals from the Uniform Crime Reporting program. They further estimate breakdowns between urban and rural agencies using individual level time diary estimates from studies specific jurisdictions, and conduct monte-carlo simulation to generate distributions around point estimates, although in this study we only use the point estimates for Part 1 crimes.

As Hunt et al. (2019) pointed out in the article, crimes that have the biggest costs often require the largest amount of manpower and time. Therefore, by reducing crime in those hot spot locations, the police department and city may be able directly to save money in reducing the burden to respond to different crimes. This may be more helpful than analyzing hot spots based on sentencing guidelines, as a department may not have as much motivation to make a change based on this data, since estimates of the benefit are not directly realized by the police department, nor will it help them justify any upfront investment in targeting such hot spots.

For an example, if a hot spot had an estimate \$1 million dollars of cost over a year, and say a police department was interested in implementing a problem-oriented approach similar to that in the Lowell hot spot study (Braga & Bond, 2008). In that study, the Lowell Police Department reduced various Part 1 crimes by around 20% to 40% relative to control areas. So if a police department can successfully replicate those same results, it would produce a savings of \$200,000 to \$400,000 dollars over a years time in resources *not spent* responding to crime in that area. One would not be able to make similar return on investment arguments to justify hot spots policing when using the prior crime harm indices based on sentencing decisions.

The crime reduction results for problem oriented policing in the Braga & Bond (2008) study are around the average effect estimates in a recent meta analysis on similar problem oriented policing interventions (Hinkle et al. 2020), and so while this does not consider the upfront cost estimate, it gives an estimate of *potential* return on investment. Clearly hot spot areas with crime costs lower than the potential intervention cost are not capable of generating a positive return on investment when only considering policing costs.

Data and Methods

Data used for the analysis all come from open data sources. Address level geocoded crime data is available from dallasopendata.com. For this analysis we use crime data from June-2014 through December-2016 as the training dataset to draw historical hot spots, and then compare to test data from January-2017 through June-2018. While additional crime data was available at the time of writing, Dallas PD underwent a transition to NIBRS reporting, which subsequently caused various anomalies in the open data (such as a dramatic drop in thefts). The Dallas open data does not contain reported rapes, so the analysis here only examines the other Part 1 UCR index crimes; murder, aggravated assault, robbery, burglary, theft, and motor-vehicle-theft. Crimes that were given the address of a Dallas PD station (or sub-station) were removed from the analysis.

For GIS files, we obtained 2017 TAAG hot spot areas from the same dallasopendata.com website. We obtained an outline of the city, as well as street centerlines, from <https://gis.dallascityhall.com/shapefileDownload.aspx>. There are a total of 54 TAAG areas. All geographic analysis was conducted using a local projection relevant for Dallas, a Lambert conformal conic projection centered in Texas.

Although the main analysis uses area as the denominator for PAI calculations, street centerlines are used as an alternative denominator in place of area for PAI calculations in supplementary analysis (Drawve & Wooditch, 2019). This supplementary analysis is available in the appendix, and largely mimics the same findings using area for PAI calculations that we focus on in the main analysis.

We eliminated several lochs from the city outline of Dallas proper, as including these areas would artificially increase PAI statistics, although they cannot reasonably have geocoded crime incidents occur within them (Wheeler & Steenbeek, 2020). Figure 1 displays a map of Dallas proper, along with the 2017 Dallas PD identified TAAG areas. While these are 2017 areas, many are historically consistent, and are the same as those reported in Worrall & Wheeler (2019).

For cost of crime estimates, we use the Texas specific estimates presented in Hunt, et al. (2019). Table 1 lists the descriptive statistics, as well as those cost of crime estimates per UCR offense. As the testing period is shorter (1.5 years vs 2.5 years), the total numbers of crimes are smaller. But on a per unit time scale, the overall crime counts between periods are largely comparable, so estimates based on the pre-data should be reasonable to forecast in time. This is especially the case since others have found micro place hot spots tend to be very stagnant over long periods of time (Andresen, Curman, & Linning, 2017; Curman, Andresen, & Brantingham, 2014; Weisburd et al., 2004; Wheeler, Worden, & McLean, 2016).

[INSERT TABLE 1 HERE AND FIGURE 1 HERE]

There are two main methodological advances presented in the paper. One is the use of the DBSCAN (**D**ensity **B**ased **SCAN**) clustering technique to create hotspots of crime (Campello et

al., 2013). Being a hierarchical clustering technique, it is well suited to identify irregular shaped hotspots, such as runs along a street (Brantingham & Brantingham, 1993; Grubestic, 2006).

While the contribution of illustrating DBSCAN is a minor change compared to other common hierarchical clustering techniques (Haberman, 2017), the main reason it was chosen is because it can easily incorporate weights into the clustering algorithm.¹ Here we use cost of crime estimates as those weights when constructing hot spots. This is similar to past work that has used crime harm estimates to construct and evaluate hot spots (Macbeth & Ariel, 2017; Ratcliffe, 2015).

For a brief description of how the clustering algorithm works, DBSCAN has two parameters that need to be chosen by the analysts; epsilon (the radius to which two points can ultimately be connected), and the minimum number of points (or here weighted points) to establish a cluster. These parameters are chosen here as 400 feet for epsilon, and \$400,000 for the total costs of crime within a hotspot. Four hundred feet was chosen as this is slightly less than the average street segment size in Dallas (Wheeler & Steenbeek, 2020). For an average length street segment in Dallas, if there were only crimes reported at either end, this technique would *not* link them together. We believe approximating a street segment length in epsilon is a reasonable distance parameter, given the large literature specifying crime hot spots as particular high crime street segments (MacBeth & Ariel, 2017; Weisburd et al., 2004; Weisburd, 2015). If a street segment only has crimes reported at two polar ends, *with no crimes in between*, we believe it is more likely they should be connected to other potential hot spots, not with each other.

Four hundred thousand dollars (in 2010 dollars) (Hunt et al., 2019) was chosen as a baseline cost of crime hot spot value for the ad-hoc reason it tended to produce a total number of

¹ This is not exclusive to DBSCAN, for example one could weight STAC clusters similarly (Block & Block, 1995). However, we do believe hierarchical clustering techniques are the most useful for irregularly shaped crime clusters.

hot spots similar to the total number of TAAG areas. \$400,000 is the minimum weight needed for the technique to return an identified cluster, it will ultimately identify areas with even more crime cost than that baseline.

For a simplified example of how DBSCAN works, imagine you have two locations, A and B, that are 300 feet apart, and each have a total of \$200,000 in crime cost. Even though each location by itself does not meet the \$400,000 threshold, combined they do. The epsilon parameter is what defines the threshold as to whether two points can be combined to consider their joint weight. Points A and B then form a *core cluster*. Points additionally within 400 feet from either A or B will additionally be considered inside the cluster, but unless all points within a particular radius of 400 feet exceed the \$400,000 in crime cost, those points will not be considered a core point for the cluster.

Thus the epsilon distance parameter and the minimum weight parameter are two details that need to be chosen by the analyst, they are not automatically chosen in an algorithmic way. This is both a strength and a weakness – there are likely no universal solutions for what those values should be to produce the ‘best’ hot spots, but that choice also allows the analyst to adjust the parameters for their own particular circumstance. So cities that are more spread out an analyst may choose a larger distance parameter, or for hot spot interventions that are more costly the analyst may make the threshold for hot spot identification larger as well. Ultimately *all* hot spot creation procedures involve some arbitrary decisions, e.g. choosing a bandwidth for a kernel density map, and the necessary decisions to use DBSCAN is no more onerous than other common hot spot techniques.

We use the open source R statistical package to conduct the analysis, in particular the *dbscan* library, with cost estimates for crime as the weights (Hahsler, Piekenbrock, Doran, 2019).

We create DBSCAN clusters based on the training dataset, and then buffer the core points within the clusters 400 feet to generate hot spot areas. We additionally merge any two clusters that this procedure results in overlaps, as well as remove any holes (areas in which it is entirely surrounded by a cluster). We then count the testing set number of crimes and their cost that fall within those hot spots, which provides an example of how such hot spots are used prospectively by police in practice (Berk, 2008). This procedure results in a total of 61 micro place clusters.

To evaluate the clusters, we then use the predictive accuracy index (PAI) to assess the accuracy of the hot spots. This common technique to measure accuracy (Chainey et al., 2008; Levine, 2008), and was used in the recent National Institute of Justice Forecasting challenge (Flaxman et al., 2019; Lee, O, & Eck, 2019; Mohler & Porter, 2018).

The PAI statistic is typically defined as the proportion of crimes captured in a hot spot area, divided by the proportion of the city the hot spot covers. For an example, if a particular set of identified hot spots covers 2% of the city, and the hot spots capture 10% of the total crime, that would result in a PAI of 20. Here we make a simple modification to the PAI to accommodate cost of crime estimates, for the numerator we consider the total cost of crime captured, instead of simply the proportion of crime. So in our prior example, if there is a total of \$1 million dollars in crime cost in a city, and the same hot spot captures \$200,000 in crime cost (20% of the total cost), using the same 2% of the area denominator, the PAI statistic would then be 40. We evaluate both the total cumulative PAI identified by our DBSCAN clusters vs Dallas PD defined TAAG areas, as well as provide graphs and maps of individual identified hot spots.

Results

Table 2 displays the resulting citywide aggregate statistics, accumulating the entire TAAG and our identified DBSCAN hot spot areas. The total crime cost captured in the

DBSCAN areas is just under \$20 million dollars, which is 12% of the cost of crime in the city (over \$169 million dollars). Given that the DBSCAN areas cover less than 3 square miles (1% of the city), this results in a PAI statistic of 16. Although the clusters were only fit on the cumulative proportions of each crime, overall they result in very similar PAI statistics when doing the (non-cost weighted) breakdowns for individual crime types, thus the results do not appear to be driven by a particular crime type.

While the TAAG hot spots capture a much larger proportion of crime and crime cost, they also cover a much larger area of the city. TAAG areas cover 65 square miles, 19% of the city. Overall TAAG's capture 54% of the costs of crime in Dallas (over \$92 million dollars), resulting in a PAI of 2.9. Again the statistics only vary slightly when considering non-weighted crime counts, and are typically around a PAI of 3.

Figure 2 displays PAI statistics for individual TAAG and DBSCAN hot spot areas. The majority of the TAAG areas hover under a PAI of 5, whereas the DBSCAN areas have a much broader distribution. While the left tail of the DBSCAN hot spots falls short of the PAI over 10 ideal laid out in the introduction, 50 of the 61 DBSCAN clusters have a PAI of over 10. It happens that only three of the DBSCAN areas have a lower PAI (based on the cost of crime) than the *highest* PAI for TAAG areas.

[INSERT TABLE 2 AND FIGURE 2 HERE]

Figure 3 displays a screenshot of an interactive map we created to explore the areas, <https://apwheele.github.io/MathPosts/HotSpotMap.html>. The blue DBSCAN areas are all smaller than 0.2 square miles, whereas the red TAAG areas tend to be around 1 to 2 square miles. Like

other work on hot spots at micro places (Weisburd et al., 2004), the DBSCAN clusters are spread throughout the city, although have several sub-clusters in different areas of the city. The majority (but not all) of DBSCAN areas are covered by TAAGs. This suggests many of the TAAG hot spots currently in place can simply be adjusted to focus on more specific places.

Another way to consider the size of a hot spot is to consider the length of the streets in the area, since that will more naturally define the area in which officers will be patrolling (Drawve & Wooditch, 2019; Ratcliffe & Sorg, 2017). For reference hot spot areas in terms of street length, in the Philadelphia policing tactics experiment, the hot spot areas tended to cover around 3 miles of streets (Groff et al., 2015). For the Philadelphia foot patrol experiment, the areas covered a total of 1.3 miles of street on average (Ratcliffe et al., 2011). The DBSCAN hot spots here on average cover just under 1 mile of street. Of the 61 DBSCAN hot spots, only 11 cover more than 1.3 miles of street length, and only one covers more than 3 miles of street length. For comparison the current TAAG areas cover on average 24 miles of streets. The smallest TAAG area in terms of street length coverage is 12 miles.

Figure 4 illustrates a more zoomed in focused screen shot of the same interactive map. The map provides tooltips when clicking on a blue DBSCAN area, and this provides a breakdown of the total crimes and the cost of crime captured in a particular DBSCAN area. The particular example shown, a cluster that runs along Scyene road and expands onto Saint Augustine road illustrates the utility of a hierarchical clustering technique. Clusters that rely on identifying elliptical shaped clusters would have a difficult time approximating that shape. This particular hot spot contains a Family Dollar store, a gas station, an elementary school, and several apartment complexes, each likely an individual high risk crime generator (Eck et al., 2007), that all in toto contribute various crimes to this hotspot. One can see from the tool tip that

the hot spot contains a variety of violent (26 aggravated assaults, 13 robberies), as well as non-violent crimes (15 burglaries, 58 thefts, and 26 thefts of motor-vehicles).

In terms of the question do hot spots tend to be a single address or a street segment, one can find examples to support either. Many of the clusters are 1 to 2 addresses, although this may in part be a function of the nature of the clustering algorithm. It can return a single address as a cluster if it happens to have more than \$400,000 dollars in cost in the training dataset. But if there are any addresses within 400 feet, even with a single crime, they will also be within the core cluster. One can also find examples that more closely approximate street segments, or several street segments nearby one another, which are a function of multiple addresses accumulating to reach that \$400,000 threshold.

[INSERT FIGURE 3 AND FIGURE 4 HERE]

Other exploratory data analysis did not reveal any obvious patterns to the identified hot spots. In particular we examined whether specific crime types tended to dominate the identified clusters. It could be since violent crimes have larger costs, the hot spots tend to be more violent on average. Or the opposite, since property crimes are much more prevalent, the hot spots are all dominated by thefts. Neither of these appeared to be the case – the majority of hot spots tended to have a wide distribution of crimes (although one could find some counter-examples dominated by violent or property crimes). To quantify this, we have estimated the Shannon entropy statistic for each of our DBSCAN clusters (Lee & Eck, 2019).² For this statistic, if a hot spot only contained one crime type, the entropy would be 0. Higher values signify more entropy, so it is

² The entropy statistic can be calculated as $E = -\sum_c p_c \cdot \log(p_c)$, where p_c is the proportion of crime c out of all crimes in that hot spot area. So for an example, a hot spot location of 5 robberies, 3 murders, and 2 motor vehicle thefts would have an entropy statistic of $E = -[\frac{5}{10} \cdot \log(\frac{5}{10}) + \frac{3}{10} \cdot \log(\frac{3}{10}) + \frac{2}{10} \cdot \log(\frac{2}{10})] \approx 1.02$. Crimes with zero values do not enter into the calculation.

harder to predict what type of crime may fall within that hot spot. The average entropy for each area is 1.38, with a maximum entropy given six potential crime types under examination here is 1.79. Thus areas are much closer to max entropy than they are encompassing a predictable set of crime types.

This is contra to the work by Haberman (2017), who found hot spots of different crime types did not tend to overlap. This difference is perhaps a result of clustering on the combined crimes, instead of clustering crimes individually. Although the difference in results may also be the case that Dallas has substantively different crime patterns than Philadelphia.

Discussion

The results of the analysis demonstrated what we believed at the onset – that Dallas PD could identify much smaller hot spots than the current TAAG areas in place and capture a much larger density of cost of crime per area. Like prior analysis of concentration of crime harm (Curtis-Ham & Walton, 2017; Fenimore, 2019; Norton et al., 2018; Weinborn et al., 2018), we find micro places that have high values of crime cost in clusters. Additionally, we show that such micro places have long term stability in the cost of crime, by showing that hot spots generated with historical data also capture high values of crime cost in the future (Norton et al., 2018). As such, our work is complimentary to many past analyses, and builds on a long standing literature of constructing and evaluating the predictive capability of hot spots clustering.

An additional finding of the work is that we identify a heterogenous set of areas that qualify as hot spots in our clustering algorithm. While much past work has either conducted analysis at a pre-specified unit of analysis, such as census areas (Curtis-Ham & Walton, 2017), street segments (Norton et al., 2018), or specific addresses (O'Brien & Winship, 2017), our approach identifies hot spots that do not nicely conform to any of those pre-specified units of

analysis. Thus there is likely a potential to be more efficient in hot spot identification when using a solution that does not presume a specific structure for hot spots at the start. Although we note that any of these ranking approaches is likely better to the status quo of constructing hot spot areas based on personal intelligence (MacBeth & Ariel, 2017).

That being said, ultimately the nature of hot spot identification will always come with some arbitrariness, given that hot spots themselves have no well agreed upon definition (Taylor, 2015). Subsequently a limitation of this work is that we do not consider here other methods used to identify hot spots and test their accuracy. Many exist, in addition to the traditional clustering techniques mentioned in the literature review, there are a variety of model based approaches (Drawve, 2016; Mohler, Carter, & Raje, 2018; Wheeler & Steenbeek, 2020) we have not touched on here. Thus while we cannot say that our DBSCAN analysis is dispositive that it is better than these other approaches, we believe it is likely the case one can use many of these different approaches to improve upon the current Dallas TAAGs, even simply just counting up street segments with the highest crimes is likely much more accurate (MacBeth & Ariel, 2017; Wheeler & Steenbeek, 2020).

One strong limitation of the current analysis is that the definition of the cost of crime is not well agreed upon. While here we use the work of Hunt et al. (2019) that is mostly based on costs associated with police response to particular crimes, there is a wide array of ways to calculate the cost of crime (Domínguez & Raphael, 2015). For example, a murder in this analysis is valued at \$124 thousand dollars, whereas the value of a statistical life is often pegged at over one million dollars (Domínguez & Raphael, 2015). This difference has to do with who bears the cost of the murder – for police it is mostly in terms of investigative time, whereas the cost is much more severe for the victims family (and infinite for the victim themselves). It may also be

that crime prevented for individuals accumulate over a lifetime (Cohen & Piquero, 2009), with failing to prevent crime resulting in long term negative externalities for various individuals. Finally, this does not consider additional considerations on allocating police resources, such as equity or preventing disproportionate minority contact with a hot spots policing approach (Wheeler, 2019). So while these cost of crimes estimates are relevant to police departments themselves, they may not be reasonable to the greater public when constructing a hot spots policing strategy.

Additionally, this work relying on public data does not include rapes, and we also do not include Part 2 crimes, which are often considered public disorder crimes and additionally contribute to quality of life (Boggess & Maskaly, 2014; Chappell, Monk-Turner, & Payne, 2010; Ratcliffe, 2015). This is not a limitation for a police department applying such estimates internally, where they have access to all of the reported crime data, but even then non-reporting will cause such estimated cost of crime hot spots to be an *underestimate* of the cost of crime at particular places. Ultimately how to appropriately value crime harm is as much an ethical question as it is an empirical. This work shows one way to do so, using costs directly relevant to police departments, but this is not to say other ways, such as via sentencing severity or public surveys are inherently inferior. But it is likely the case that many of these different ranking schemes will produce similar rankings in crime (e.g. all would likely identify homicide as a higher harm than theft), so we believe using *any* crime harm ranking scheme is likely better than the defacto standard of treating such crimes equally when creating hot spots.

So we believe that both of these biases likely underestimate the cost of crime that is occurring the hot spots we identify here. Given that the ultimate goal of identifying such hot spots is to prevent crime, these estimates may then be gross underestimates of the potential return

on investment if Dallas PD targets crime in those identified areas. But that is also conditional on a hot spots policing strategy to be effective (Mitchell, 2019) – just placing hot spots on the map and not doing anything with that information will ultimately not reduce crime at those locations.

Related to this is that in the DBSCAN algorithm itself we need to define an agreed upon *minimal threshold* at which to identify a location is a hot spot. Here we arbitrarily choose \$400,000 dollars. Ultimately this minimum threshold would be better determined by the nature of the intervention police departments wish to engage in that hot spot area. So for example if the particular intervention was more cost intensive (e.g. overtime patrols), this is would likely only be justified given higher cost thresholds. In that case, a \$400,000 hot spot over a year may not be sufficient, and the department may wish to raise this threshold. Cheaper interventions, such as nudging officers to do more patrols in high crime areas (Mohler et al., 2015), may however justify lowering the threshold to lower dollar amounts.

While this work focuses on long term crime forecasting, the application of using DBSCAN can potentially be extended to short term crime forecasting (Flaxman et al., 2019; Garnier, Caplan, & Kennedy, 2018; Lee et al., 2019; Mohler et al. 2015; Ratcliffe et al., 2020). One approach that we believe may be fruitful is to use DBSCAN on the output of a predictive policing application applied at the *address and intersection* level to create homogenous areas to assign subsequent patrols (Deryol et al., 2016). Although raster based approaches to predictive policing are popular (Caplan & Kennedy, 2011), these raster grids do not conform to the actual micro places and the street network that heavily influence crime (Groff, 2014; Rosser et al., 2017). While prior work has used the creation of the raster grid orientation and size of the grid cells as a hyperparameter when tuning machine learning models (Flaxman et al., 2019; Mohler & Porter, 2018), one could similarly use the DBSCAN parameters of epsilon and minimum weight

in the same capacity to avoid relying in grid cells entirely. Such results are likely to result in more natural, contiguous hot spot areas to target. Such contiguous areas would not only require fewer resources to target (relative to separate hot spots not in the same area), but also would likely result in better police adherence when targeting those hot spots boundaries (Sorg et al., 2017).

Given that errors associated with geocoding tend to be around 100 feet on average for crimes occurring outdoors (Wheeler, Gerell, & Yoo, 2020), it likely does not make sense to cluster crimes within a smaller distance than that shown here, but it may be the case that one would identify more agglomerated areas if the epsilon clustering criteria was slightly larger, as one can see sub-clusters of the individual clustered hot spot areas in the resulting map in this analysis. It is likely the case that idiosyncratic characteristics of different jurisdictions will likely change what distance parameter captures the most weighted crime cost, so future research may be warranted about how to best tune this value given a similar test and validation data set.

Policy Implications

The policy implications of constructing hot spots based on cost-of-crime estimates, as opposed to either direct counts of crime or weights based on sentencing severity are rather straightforward. They provide a much more direct cost-benefit analysis of the potential return on investment directly relevant to police departments.

For a direct estimate, let us take an example from the hot spot that captured the largest cost of crime, a hot spot in the center of Dallas that captured around 1.5 million dollars of crime within 1.5 years. The average crime reduction in the Braga et al. (2019) meta-analysis is around 20%. As such, the expected return to partaking in a hot spots policing strategy at this one location is around \$200,000 per year. Such a return on investment would justify assigning

additional resources to target this hot spot. For locations with smaller overall costs of crime, while it likely does not justify allocating additional resources, it certainly may justify reallocating existing resources to target crime in those particular micro places.

While this analysis is illustrative of how one might create cost-of-crime weighted hot spots, ultimately the optimal strategy to construct hot spots needs to entail what actions the police department is going to take within those hot spots (as well as functional constraints in the resources the police department can allocate). Historically Dallas PD has only generically stated TAAG area hot spots were formulated to predict areas of high crime (Ferguson, 2017). They have not articulated specific strategies they would undertake at said hot spots.

Take for example if Dallas PD knew they wanted to create foot or bike patrols for identified hot spots, and budgeted enough to assign 4 different officers on some regular basis. It is likely the case that only four officers cannot cover all 61 micro place hot spots identified in this analysis, but it would be easily possible for those officers to cover at least one of the micro place hot spots identified here, which are all under 0.2 square miles (Haberman & Stiver, 2019; Ratcliffe & Sorg, 2017). Given that such areas have a specific cost attached to them, it also provides more straightforward cost-benefit analysis if the Dallas PD wishes to advocate for more resources to be able to cover such hot spots, or to justify reallocating existing resources to cover such hot spots.

While those same officers could cover the existing TAAG areas, expanding the size of such hot spots will result in a more diffused intervention by the police, no matter what the intervention is (Larson, 1975). As such it is likely in Dallas PD's best interest to take a more vested approach to identify micro places, even if they do not have a specific plan for what to do in the current hot spots at the current time. It may be a police department wishes to generate such

hot spots first, and then take a problem oriented approach to solving crimes in those areas (Braga et al., 1999). Given that targeted police interventions at large areas tend to be less successful than micro placed focused interventions (Lum, Koper, & Telep, 2011), weighted cost-of-crime DBSCAN identified hot spots are a tool crime analysts for any police department should consider in formulating a hot spots policing strategy.

The feasibility of such an approach should be within the capabilities of Dallas PD to implement. This analysis comes with a set of replication materials using open source data and open source code, thus Dallas PD (or any other interested police department) can replicate the analysis using more current data for free if they so wish. Given the consistency of other papers identifying hot spots of crime, it is likely the case such work can generate areas of hot spots which show promise of generating significant returns on investment for police departments. Hot spots weighted by cost of crime estimates can then be used as an upfront tool to justify either new investments in PDs to allocate resources to hot spots or to shift current resources to hot spots. This provides a direct cost-benefit calculation to justify to police departments to shift resources from reactive to proactive policies (Nagin et al., 2015).

References

- Andresen MA, Curman AS, Linning SJ (2017) The trajectories of crime at places: Understanding the patterns of disaggregated crime types. *Journal of Quantitative Criminology* 33: 427-449.
- Bayley DH (1994) *Police for the Future*. Studies in Crime and Public Policy.
- Berk R (2008) Forecasting methods in crime and justice. *Annual Review of Law and Social Sciences* 4: 219-238.
- Block RL, Block CR (1995) Space, place and crime: Hot spot areas and hot places of liquor-related crime. *Crime Prevention Studies* 4: 145-184.

- Boggess L, Maskaly J (2014) The spatial context of the disorder-crime relationship in a study of Reno neighborhoods. *Social Science Research* 43: 168-183.
- Braga AA, Bond BJ (2008) Policing crime and disorder hot spots: A randomized controlled trial. *Criminology* 46: 577-607.
- Braga AA, Papachristos AV, Hureau DM (2010) The concentration and stability of gun violence at micro places in Boston, 1980-2008. *Journal of Quantitative Criminology* 26: 33-53.
- Braga AA, Turchan BS, Papachristos AV, Hureau DM (2019) Hot spots policing and crime reduction: An update of an ongoing systematic review and meta-analysis. *Journal of Experimental Criminology* 15: 289-311.
- Braga AA, Weisburd DL, Waring EL, Mazerolle LG, Spelman W, Gajewski F (1999) Problem-oriented policing in violent crime places: A randomized controlled experiment. *Criminology* 37: 541-580.
- Brantingham PL, Brantingham PJ (1993) Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology* 13: 3-28.
- Campello, RJGB, Moulavi D, Sander J (2013) Density-Based Clustering Based on Hierarchical Density Estimates. *Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery in Databases, PAKDD 2013, Lecture Notes in Computer Science* 7819: 160.
- Caplan JM, Kennedy LW (2011) Risk terrain modelling: Brokering criminological theory and GIS methods for crime forecasting. *Justice Quarterly* 28: 360-381.
- Chainey S, Thompson L, Uhlig S (2008) The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal* 21: 4-28.
- Chappell A, Monk-Turner E, Payne B (2010). Broken windows or window breakers: The influence of physical and social disorder on quality of life. *Justice Quarterly* 28: 522-540.
- Cohen MA, Piquero AR (2009) New evidence on the monetary value of saving a high risk youth. *Journal of Quantitative Criminology* 25: 25-49.
- Curman, AS, Andresen MA, Brantingham PJ (2014) Crime and place: A longitudinal examination of street segment patterns in Vancouver, BC. *Journal of Quantitative Criminology* 31: 127-147.
- Curtis-Ham S, Walton D (2017) Mapping crime harm and priority locations in New Zealand: A comparison of spatial analysis methods. *Applied Geography* 86: 245-254.

- Curtis-Ham S, Walton D (2018) The New Zealand crime harm index: Quantifying harm using sentencing data. *Policing: A Journal of Policy and Practice* 12: 455-467.
- Deryol R, Wilcox P, Logan M, Wooldredge J (2016) Crime places in context: An illustration of the multilevel nature of hotspot development. *Journal of Quantitative Criminology* 32: 305-325.
- Domínguez P, & Raphael S (2015) The role of the cost-of-crime literature in bridging the gap between social science research and policy making: Potentials and limitations. *Criminology & Public Policy* 14: 589-632.
- Drawve G (2016) A metric comparison of predictive hot spot techniques and RTM. *Justice Quarterly* 33: 369-397.
- Drawve G, Wooditch A (2019) A research note on the methodological and theoretical considerations for assessing crime forecasting accuracy with the predictive accuracy index. *Journal of Criminal Justice* Online First.
- Eck JE, Chainey SP, Cameron JG, Leitner M, Wilson RE (2005) *Mapping Crime: Understanding Hot Spots*. USA: National Institute of Justice.
- Eck JE, Clarke RV, Guerette RT (2007) Risky facilities: Crime concentration in homogenous setes of establishments and facilities. *Crime Prevention Studies* 21: 225-264.
- Fenimore DM (2019) Mapping harmspots: An exploration of the spatial distribution of crime harm. *Applied Geography* 109: 102034.
- Ferguson AG (2017) *The rise of big data policing: Surveillance, race, and the future of law enforcement*. New York: New York University Press
- Flaxman S, Chirico M, Pereira P, Loeffler C (2019) Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the NIJ “real-time crime forecasting challenge”. *The Annals of Applied Statistics* 13: 2564-2585.
- Garnier S, Caplan JM, Kennedy LW (2018) Predicting dynamical crime distribution from environmental and social influences. *Frontiers in Applied Mathematics and Statistics* 4: 13.
- Groff ER (2014) Quantifying the exposure of street segments to drinking places nearby. *Journal of Quantitative Criminology* 30: 527-548.

- Groff ER, Ratcliffe JH, Haberman CP, Sorg ET, Joyce NM, Taylor RB (2015) Does what police do at hot spots matter? The Philadelphia policing tactics experiment. *Criminology* 53: 23-53.
- Grubestic TH (2006) On the application of fuzzy clustering for crime hot spot detection. *Journal of Quantitative Criminology* 22: 77.
- Haberman CP (2017) Overlapping hot spots? Examination of the spatial heterogeneity of hot spots of different crime types. *Criminology & Public Policy* 16: 633-660.
- Haberman CP, Stiver WH (2019) The Dayton foot patrol program: an evaluation of hot spots foot patrols in a central business district. *Police Quarterly* 22: 247-277.
- Hahsler M, Piekenbrock M, Doran D (2019) dbscan: Fast Density-Based Clustering with R. *Journal of Statistical Software* 91: 1-30.
- Hinkle JC, Weisburd D, Telep CW, Petersen K (2020) Problem-oriented policing reducing crime and disorder: An updated systematic review and meta-analysis. *Campbell Systematic Reviews* 16: 2.
- House PD, Neyroud PW (2018) Developing a crime harm index for Western Australia: The WACHI. *Cambridge Journal of Evidence-Based Policing* 2: 70-94.
- Hunt PE, Anderson J, Saunders J (2017) The price of justice: new national and state-level estimates of the judicial and legal costs of crime to taxpayers. *American Journal of Criminal Justice* 42: 231-254.
- Hunt PE, Saunders J, Kilmer B (2019) Estimates of law enforcement costs by crime type for benefit-cost analyses. *Journal of Benefit-Cost Analysis* 10: 95-123.
- Ignatans D, Pease K (2016) Taking crime seriously: playing the weighting game. *Policing: A Journal of Policy and Practice* 10: 184-193.
- Kounadi O, Ristea A, Araujo Jr A, Leitner M (2020) A systematic review on spatial crime forecasting. *Crime Science* 9: 7.
- Larson RC (1975) What happened to patrol operations in Kansas City? A review of the Kansas City preventive patrol experiment. *Journal of Criminal Justice* 3: 267-297.
- Lee Y, Eck JE (2019) Comparing measures of the concentration of crime at places. *Crime Prevention and Community Safety* 21: 269-294.

- Lee YJ, O SH, Eck JE (2019) A theory driven algorithms for real-time crime hot spot forecasting. *Police Quarterly* Online First.
- Levine N (2008) The “Hottest” part of a hotspot: Commentary on “The utility of hotspot mapping for predicting spatial patterns of crime”. *Security Journal* 21: 295-302.
- Lum C, Koper CS, Telep CW (2011) The evidence-based policing matrix. *Journal of Experimental Criminology* 7: 3-26.
- Macbeth E, Ariel B (2017) Place-based statistical versus clinical predictions of crime hot spots and harm locations in Northern Ireland. *Justice Quarterly* 36: 93-126.
- Mitchell RJ (2019) The usefulness of a crime harm index: analyzing the Sacramento Hot Spot Experiment using the California Crime Harm Index (CA-CHI). *Journal of Experimental Criminology* 15: 103-113.
- Mohler GO, Carter J, Raje R (2018) Improving social harm indices with a modulated Hawkes process. *International Journal of Forecasting* 34: 431-439.
- Mohler GO, Porter MD (2018) Rotational grid, PAI-maximizing crime forecasts. *Statistical Analysis and Data Mining* 11: 227-236.
- Mohler GO, Short MB, Malinowski S., Johnson M., Tita GE, Bertozzi AL, Brantingham PJ (2015) Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association* 110:1399-1411.
- Nagin DS, Solow RM, Lum C (2015) Deterrence, criminal opportunities, and police. *Criminology* 53: 74-100.
- Norton S, Ariel B, Weinborn C, O'Dwyer E (2018) Spatiotemporal patterns and distributions of harm within street segments. *Policing: An International Journal* 41: 352-371.
- O'Brien DT, Winship C (2017) The gains of greater granularity: The presence and persistence of problem properties in urban neighborhoods. *Journal of Quantitative Criminology* 33: 649-674.
- Ratcliffe, JH (2015) Towards an index for harm-focused policing. *Policing: A Journal of Policy and Practice* 9: 164-182.
- Ratcliffe JH, Sorg ET (2017) *Foot patrol: rethinking the cornerstone of policing*. Springer.

- Ratcliffe JH, Taylor RB, Perenzin-Askey A, Thomas K, Grasso J, Bethel K, Fisher R, Koehnlein J (2020) The Philadelphia predictive policing experiment. *Journal of Experimental Criminology* Online First.
- Ratcliffe JH, Taniguchi T, Groff ER, Wood JD (2011) The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology* 49: 795-831.
- Rosser G, Davies T, Bowers KJ, Johnson SD, Cheng T (2017) Predictive crime mapping: Arbitrary grids or street networks? *Journal of Quantitative Criminology* 33: 569-594.
- Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology* 27: 27-56
- Sherman LW, Neyroud PW, 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: 171-183.
- Sherman LW, Cambridge University associates (2020) How to count crime: The Cambridge Harm Index consensus. *Cambridge Journal of Evidence-Based Policing* 4: 1-14.
- Sorg ET, Wood JD, Groff ER, Ratcliffe JH (2017) Explaining dosage diffusion during hot spot patrols: An application of optimal foraging theory to police officer behavior. *Justice Quarterly* 34: 1044-1068.
- Taylor RB (2015) *Community criminology: Fundamentals of spatial and temporal scaling, ecological indicators, and selectivity bias*. NYU Press.
- Van Patten I, McKeldin-Conor J, Cox D (2009) A microspatial analysis of robbery: Prospective hot spotting in a small city. *Crime Mapping: A Journal of Research and Practice* 1: 7-32.
- Weinborn C, Ariel B, Sherman LW, O'Dwyer E (2017) Hotspots vs. harmspots: Shifting the focus from counts to harm in the criminology of place. *Applied Geography* 86: 226-244.
- Weisburd D (2015) The law of crime concentration and the criminology of place. *Criminology* 53: 133-157.
- Weisburd DL, Bushway SD, Lum C, Yang, SM (2004) Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology* 42: 283-322
- Wheeler AP (2019) Allocating police resources while limiting racial inequality. *Justice Quarterly* Online First.

Wheeler A, Gerell M, Yoo Y (2019) Testing the Spatial Accuracy of Address Based Geocoding for Gun Shot Locations. *The Professional Geographer* Online First.

Wheeler AP, Steenbeek W (2020) Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology* Online First.

Wheeler AP, Worden RE, McLean SJ (2016) Replicating group-based trajectory models of crime at micro places in Albany, NY. *Journal of Quantitative Criminology* 32: 589-612.

Wolfgang M (1985) *The National Survey of Crime Severity*. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Worrall JL, Wheeler AP (2019) Evaluating Community Prosecution Code Enforcement in Dallas, Texas. *Justice Quarterly* 36: 870-899.

Figure 1: Historical hot spots used by the Dallas Police Department, Target Area Action Grid (TAAG) areas.

Figure 2: PAI statistics for individual DBSCAN crime harm weighted hotspots vs current TAAG areas.

Figure 3: Screenshot of interactive map showing superimposed TAAG areas (red) with the constructed DBSCAN hot spot areas (Blue). Map available at <https://apwheelee.github.io/MathPosts/HotSpotMap.html>.

Figure 4: Zoomed in screenshot of interactive map showing superimposed TAAG areas (red) with the constructed DBSCAN hot spot areas (Blue). Blue DBSCAN areas have a textual description showing crime totals and harms totals for the individual area. Map available at <https://apwheelee.github.io/MathPosts/HotSpotMap.html>.

Table 1: Cost of Crime and Crime Totals

Crime	Cost of Crime	Training Totals	Test Totals
Aggravated Assault	8,292	17,485	9,864
Burglary	1,185	27,928	12,415
Murder	124,353	299	179
Robbery	2,229	9,879	4,932
Theft	1,024	63,159	31,134
Motor Vehicle Theft	769	18,209	10,029

Table 2: Cumulative PAI statistics for DBSCAN areas vs TAAG areas

	DBSCAN	% Tot. DBSCAN	PAI DBSCAN	TAAG	% Tot. TAAG	PAI TAAG	Citywide
Total Harm (in Thousands)	19,815	12%	16.0	92,155	54%	2.9	169,350
Aggravated Assault	1,414	14%	19.6	5,405	55%	2.9	9,864
Burglary	924	7%	10.2	6,071	49%	2.6	12,415
Murder	20	11%	15.3	106	59%	3.1	179
Robbery	631	13%	17.5	3,018	61%	3.2	4,932
Theft	2,463	8%	10.8	15,814	51%	2.7	31,134
Motor Vehicle Theft	753	8%	10.3	5,254	52%	2.8	10,029
Area (Square Miles)	2.5	1%		65.0	19%		342

Appendix: PAI Statistics using Street Length instead of Area

This appendix provides supplementary results when calculating PAI statistics using the street length contained within a hotspot instead of the area (Drawve & Wooditch, 2019). They show comparable results. The cumulative citywide statistics show a smaller PAI totals for both the DBSCAN and TAAG areas, falling short of the PAI 10 ideal. However, when viewing individual hot spots identified, many have quite high weighted PAI statistics, and many of the TAAG areas overall are quite low.

Figure A.1: PAI Statistics for Individual Hot Spot Areas, using the percent of weighted crime harm as the numerator, and the percent street length within a hot spot as the denominator.

Table A.1: Cumulative Hot Spot Statistics, comparing PAI using area vs Street Length

	DBSCAN	% Tot DBSCAN	PAI DBSCAN (Area)	PAI DBSCAN (Len.)	TAAG	% Tot TAAG	PAI TAAG (Area)	PAI TAAG (Len)	Total
Total Harm (in Thousands)	19,815	12%	16.0	10.6	92,155	54%	2.9	2.3	169,350
Aggravated Assault	1,414	14%	19.6	13.0	5,405	55%	2.9	2.3	9,864
Burglary	924	7%	10.2	6.8	6,071	49%	2.6	2.1	12,415
Murder	20	11%	15.3	10.1	106	59%	3.1	2.5	179
Robbery	631	13%	17.5	11.6	3,018	61%	3.2	2.6	4,932
Theft	2,463	8%	10.8	7.2	15,814	51%	2.7	2.2	31,134
Motor Vehicle Theft	753	8%	10.3	6.8	5,254	52%	2.8	2.2	10,029
Area (Square Miles)	2.5	1%			65.0	19%			342
Length Streets (Miles)	60.6	1%			1,288.1	23%			5,498

Thank you Dr. Worrall and the three reviewers for the feedback. Below is my response to each point-by-point (my responses highlighted in grey following original reviewer text unedited).

Reviewer: 1

In terms of methodology, I believe that this paper and method does contribute to the literature. My concerns center on theoretical and conceptual aspects of the paper as I point out below. The authors might consider these points in revising the manuscript.

1. Page 2: Minor point, but there is an updated Braga et al. systematic (2019) review of Hot Spots Policing

Response: Done – I have cited the Journal of Experimental Criminology article (appears to be redundant with the Campbell Systematic paper).

2. Page 6, first paragraph under “Measuring hot spot accuracy” heading. After reading this paragraph I was reminded of a somewhat obscure paper that Ralph Taylor wrote several years ago. See: "Hot spots do not exist, and four other fundamental concerns about hot spots policing." This would be an oversimplification of Taylor's main points, but an important takeaway is that hot spots exist in the data world, but not the real world. Therefore, it is impossible to gauge the accuracy of hot spotting techniques. This shouldn't be construed as a criticism of this work, but rather this literature as a whole. I think it would be worth including a discussion of this in the paper and revising this section generally to either point out or rebut his commentary.

Response: Yes, Taylor makes the same point in his book that I do cite. Note the prior section was titled 'Creating Hot Spots'. I have expanded this paragraph in the prior section (first sentence was already in the manuscript):

However, despite the popularity of hot spots policing, there is not a single, unanimous definition of what constitutes a hot spot of crime (Taylor, 2015). This is because a hot spot is ultimately a data based definition, not a physical entity that one can go and show in the physical world. Thus while it is easy to show that a small number of micro locations have a relatively large number of crime counts (Weisburd, 2015), it is much harder to delineate the boundaries of what is inside or what is outside of a hot spot (Taylor, 2015). Subsequently there have been a myriad of different ways using data to define hot spots.

It isn't impossible to gauge the accuracy of hot spotting techniques (indeed this work illustrates the opposite), it is *only necessary* because there are not uniform definitions of what a hot spot is. I have added this paragraph into the discussion

That being said, ultimately the nature of hot spot identification will always come with some arbitrariness, given that hot spots themselves have no well agreed upon definition (Taylor, 2015). Subsequently a limitation of this work is that we do not consider here other methods used to identify hot spots and test their accuracy. Many exist, in addition to the traditional clustering techniques mentioned in the literature review, there are a variety of model based approaches (Drawve, 2016; Mohler, Carter, & Raje, 2018; Wheeler & Steenbeek, 2020) we have not touched

on here. Thus while we cannot say that our DBSCAN analysis is dispositive that it is better than these other approaches, we believe it is likely the case one can use many of these different approaches to improve upon the current Dallas TAAGs, even simply just counting up street segments with the highest crimes is likely much more accurate (MacBeth & Ariel, 2017; Wheeler & Steenbeek, 2020).

3. Pages 10-11: Jerry Ratcliffe and others have written about operationalizing harm (the authors do discuss and cite this literature), and as we know it harm refers to harms being experienced by the community. Here, the authors refer to harm, but they are referring to harm experienced by the police department in terms of cost. I would be careful about equating this technique as addressing “harm” or being a type of harm score and discussing it in terms of harm to police. I would levy the same criticism at the paper they cite to justify this decision. The two are not the same. I think this discussion needs to be re-framed, perhaps in the context of the budget tightening that many departments are likely to face moving forward given the current sentiment toward policing in the wake of the death of George Floyd. I recognize that this is a methodological decision for the authors, but it is conceptually problematic.

Response: I have changed the wording in much of the manuscript when it is in reference to our own analysis to replace ‘crime harm’ with ‘crime cost’. I have added this to the discussion section as well (and the prior two paragraphs already discussed how valuating harm is potentially arbitrary).

Ultimately how to appropriately value crime harm is as much an ethical question as it is an empirical. This work shows one way to do so, using costs directly relevant to police departments, but this is not to say other ways, such as via sentencing severity or public surveys are inherently inferior. But it is likely the case that many of these different ranking schemes will produce similar rankings in crime (e.g. all would likely identify homicide as a higher harm than theft), so we believe using *any* crime harm ranking scheme is likely better than the defacto standard of treating such crimes equally when creating hot spots.

I have not added anything in particular to tightening police budgets (which is a perpetual argument, e.g. austerity cuts in UK, not just a recent phenomenon). My arguments stand whether that occurs or not – my identified cost of crime hot spots have potential to generate appreciable returns to investment on conducting a hot spots policing strategy.

4. Page 12: Although the authors use previous estimates of the cost of crime and point readers to that literature, I think a fuller discussion of how these estimates were derived would be useful here. Simply identifying an estimate from previous literature without a discussion of how these figures were arrived at seems like a gap given that the authors identify inclusion of a cost-estimate in their hot spotting technique as a contribution.

Response: I have added in this section to further detail the Hunt methodology.

To briefly describe Hunt et al.'s (2019) methodology in more detail, it specifically focuses on the labor costs of *responding* to crime. Criminologists often distinguish between reactive and proactive policing (Nagin et al., 2015), and Hunt's estimates are relevant for the reactive category. The reactive police response to crime can be broken down in-between various categories; Hunt et al. (2019) list different categories administrative, arrest, crime scene, court

time, investigative, and en-route/waiting. They generate state level estimates for costs of responding to crime based on various data sources, including Bureau of Justice Statistics resources for justice expenditures and time spent responding to different crimes (broken down by differing roles, such as patrol officer vs detective), and state crime incident totals from the Uniform Crime Reporting program. They further estimate breakdowns between urban and rural agencies using individual level time diary estimates from studies specific jurisdictions, and conduct monte-carlo simulation to generate distributions around point estimates, although in this study we only use the point estimates for Part 1 crimes.

5. Page 13: The authors point out the arbitrariness of certain hot spot identification techniques in previous parts of the paper, but in deciding upon a 400 foot parameter they are making an arbitrary decision as well. This is true despite the fact that it has precedent in the literature. I think a theoretical, not methodological, discussion for why 400 feet is appropriate is needed here.

Response: I have added this into that section to justify this decision.

For an average length street segment in Dallas, if there were only crimes reported at either end, this technique would *not* link them together. We believe approximating a street segment length in epsilon is a reasonable distance parameter, given the large literature specifying crime hot spots as particular high crime street segments (MacBeth & Ariel, 2017; Weisburd et al., 2004; Weisburd, 2015). If a street segment only has crimes reported at two polar ends, *with no crimes in between*, we believe it is more likely they should be connected to other potential hot spots, not with each other.

I have also edited the last sentence in this paragraph in the Discussion:

Given that errors associated with geocoding tend to be around 100 feet on average for crimes occurring outdoors (Wheeler, Gerell, & Yoo, 2020), it likely does not make sense to cluster crimes within a smaller distance than that shown here, but it may be the case that one would identify more agglomerated areas if the epsilon clustering criteria was slightly larger, as one can see sub-clusters of the individual clustered hot spot areas in the resulting map in this analysis. It is likely the case that idiosyncratic characteristics of different jurisdictions will likely change what distance parameter captures the most weighted crime cost, so future research may be warranted about how to best tune this value given a similar test and validation data set.

6. Page 14: \$400k is another arbitrary decision. Although it might be useful in the paper for comparison of TAAG, what does this mean in terms of applicability in real world policing? The discussion beginning on page 17, line 40 highlights the need to get the parameters right.

Response: I have added this paragraph in the Discussion section in response to this:

Related to this is that in the DBSCAN algorithm itself we need to define an agreed upon *minimal threshold* at which to identify a location is a hot spot. Here we arbitrarily choose \$400,000 dollars. Ultimately this minimum threshold would be better determined by the nature of the intervention police departments wish to engage in that hot spot area. So for example if the particular intervention was more cost intensive (e.g. overtime patrols), this would likely only be

justified given higher cost thresholds. In that case, a \$400,000 hot spot over a year may not be sufficient, and the department may wish to raise this threshold. Cheaper interventions, such as nudging officers to do more patrols in high crime areas (Mohler et al., 2015), may however justify lowering the threshold to lower dollar amounts.

7. Discussion section: Again, I think this needs to be reframed in light of what was raised in point 3 above.

Response: Again, the discussion section already had an extensive discussion of this. I copy those paragraphs here again for everyone's reference. All of the historical crime weighting schemes are subject to this same criticism, since it will ultimately be arbitrary what gets counted in crime harm. I suspect many of these systems however will be correlated.

One strong limitation of the current analysis is that the definition of the cost of crime is not well agreed upon. While here we use the work of Hunt et al. (2019) that is mostly based on costs associated with police response to particular crimes, there is a wide array of ways to calculate the cost of crime (Domínguez & Raphael, 2015). For example, a murder in this analysis is valued at \$124 thousand dollars, whereas the value of a statistical life is often pegged at over one million dollars (Domínguez & Raphael, 2015). This difference has to do with who bears the cost of the murder – for police it is mostly in terms of investigative time, whereas the cost is much more severe for the victims family (and infinite for the victim themselves). It may also be that crime prevented for individuals accumulate over a lifetime (Cohen & Piquero, 2009), with failing to prevent crime resulting in long term negative externalities for various individuals. Finally, this does not consider additional considerations on allocating police resources, such as equity or preventing disproportionate minority contact with a hot spots policing approach (Wheeler, 2019). So while these cost of crimes estimates are relevant to police departments themselves, they may not be reasonable to the greater public when constructing a hot spots policing strategy.

Additionally this work relying on public data does not include rapes, and we also do not include Part 2 crimes, which are often considered public disorder crimes and additionally contribute to quality of life (Boggess & Maskaly, 2014; Chappell, Monk-Turner, & Payne, 2010; Ratcliffe, 2015). This is not a limitation for a police department applying such estimates internally, where they have access to all of the reported crime data, but even then non-reporting will cause such estimated cost of crime hot spots to be an *underestimate* of the cost of crime at particular places. Ultimately how to appropriately value crime harm is as much an ethical question as it is an empirical. This work shows one way to do so, using costs directly relevant to police departments, but this is not to say other ways, such as via sentencing severity or public surveys are inherently inferior. But it is likely the case that many of these different ranking schemes will produce similar rankings in crime (e.g. all would likely identify homicide as a higher harm than theft), so we believe using *any* crime harm ranking scheme is likely better than the defacto standard of treating such crimes equally when creating hot spots.

So we believe that both of these biases likely underestimate the cost of crime that is occurring the hot spots we identify here. Given that the ultimate goal of identifying such hot spots is to prevent crime, these estimates may then be gross underestimates of the potential return on investment if Dallas PD targets crime in those identified areas. But that is also conditional on

a hot spots policing strategy to be effective (Mitchell, 2019) – just placing hot spots on the map and not doing anything with that information will ultimately not reduce crime at those locations.

Reviewer: 2

I reviewed the manuscript titled, “Redrawing hot spots of crime in Dallas, Texas.” Overall, the paper read well. The comments below are for the author(s) should they choose to revise as needed.

Response: Thank you. As a note, Reviewer 2 suggests several additional analyses. We do not mind editing language to be clearer, or conducting a supplemental analysis that only takes a short time (using the data we have already compiled). Several of the suggestions, such as to create our own cost of crime estimates using NIBRS, and creating our own DBSCAN algorithm to account for differences in the street grid layout, would be quite an undertaking, and are unreasonable expectations to do for a paper revision.

Abstract:

-The abstract reads technical overall. I appreciate the detail within it, but the lay reader could be dissuaded from reading based on concepts used within the abstract.

Response: I’ve deleted the word ‘hierarchical’, and the note about the predictive accuracy index. It now stands as:

In this work we evaluate the predictive capability of identifying long term, micro place hot spots in Dallas, Texas. We create hot spots using a clustering algorithm, using law enforcement cost of responding to crime estimates as weights. Relative to the much larger current hot spot areas defined by the Dallas Police Department, our identified hot spots are much smaller (under 3 square miles), and capture crime cost at a higher density. We also show that the clustering algorithm captures a wide array of hot spot types; some one or two addresses, some street segments, and others an agglomeration of larger areas. This suggests identifying hot spots based on a specific unit of aggregation (e.g. addresses, street segments), may be less efficient than using a clustering technique in practice.

Introduction:

-Pg 2: Crime harm is not a well-known concept. The author(s) should consider providing the reader greater detail here

Response: I have added these few sentences.

Crime harm has been characterized in prior research by either asking survey respondents to rank different crimes (Wolfgang, 1985), or by translating sentencing decisions to create harm weights (Ratcliffe, 2015), with the ultimate goal that policing resources are better allocated relative to the harm a particular crime impacts on the community, as opposed to counting all crime equal (Sherman & Cambridge University Associates, 2020).

-Pg 3: Do the author(s) have details on how Dallas PD identifies their TAAG? The 20-60 seems like an outdated approach at face value (also providing reasoning for the current study)

Response: I don't – I have asked various individuals and have never received a specific response. Nothing official besides very vague press releases now and then as well. So I cannot personally say beyond that they are intended to identify high crime areas in Dallas.

-Pg 4: Interesting adaptation of the PAI – makes sense

Response: Thank you!

Literature Review:

-Pg 5: The placement of the Haberman paragraph should be moved to assist in flow. The paragraph has relevant information but could be better suited at the end of that subsection as an, Overall.....

Response: We have switched this placement as suggested and changed the transition to the following section.

-Pg 8: At one point, Dallas had their NIBRS data posted on their data portal – Would this be better suited for a crime harm score approach given the details captured surrounding victims and incidents (differing from UCR)?

Response: This is not a bad idea, but would require a mapping from NIBRS to a cost, which would be quite an undertaking. Honestly did not use the NIBRS data because they have some funny business going on in the reporting on the public data portal (some dramatic drops in certain crime categories), so cut off the analysis before that time point.

-Pg 9: The author(s) should consider summarizing the different harm indices by country into one overarching paragraph

Response: I have added a sign post at the intro to this section instead of a summary at the end:

Many researchers have begun developing additional harm indexes to more appropriately measure crime than by a count or rate within an area. The majority of these harm indexes are based on sentencing data, as opposed to using public opinion surveys. Although the methods to create such harm indexes have varied slightly over different locations.

-Pg 11: The author(s) should include the new meta-analysis on POP for citation support

Response: I have added these few sentences after discussing the Braga/Bond example.

The crime reduction results for problem oriented policing in the Braga & Bond (2008) study are around the average effect estimates in a recent meta analysis on similar problem oriented policing interventions (Hinkle et al. 2020), and so while this does not consider the upfront cost estimate, it gives an estimate of *potential* return on investment. Clearly hot spot areas with crime costs lower than the potential intervention cost are not capable of generating a positive return on investment when only considering policing costs.

Data and Methods:

-Pg 12: Do the author(s) have any insights on Dallas PD's analytical capabilities? Have crime analyst(s) – could they replicate, make part of their practice the approach outlined in the current study? Some insights into the PD could be helpful for a practitioner audience to translate this to be feasible

Response: I don't know exactly what you want here. From personal experience with DPD, they are a shop highly focused on intelligence, and do very little number crunching (which likely led to the TAAG definition being so out of line with best crime analyst practices).

I have provided replication code using open source software. If PDs do not have the capabilities to replicate the work (which many admittedly don't), they need to invest in competent enough data analysts to be able to conduct such analysis.

-Figure 1: Do the author(s) have access to land-use data? Might be beyond the scope of the current analysis but would be interesting to control for land-use and calculate crime harm PAI

Response: This suggestion does not make sense, and confuses the nature of the analysis. The crime hot spots are not generated based on a regression model that incorporates land use. How exactly would one 'control for land-use' in DBSCAN (an unsupervised clustering technique)? Or how would one 'control for' different land use characteristics when evaluating predictive accuracy?

I did provide a description of how one could use DBSCAN on the backend of model based estimates like RTM, copied below, but I have no idea how one would use them to evaluate the predictive accuracy directly as the reviewer suggests here.

While this work focuses on long term crime forecasting, the application of using DBSCAN can potentially be extended to short term crime forecasting (Flaxman et al., 2019; Garnier, Caplan, & Kennedy, 2018; Lee et al., 2019; Mohler et al. 2015; Ratcliffe et al., 2020). One approach that we believe may be fruitful is to use DBSCAN on the output of a predictive policing application applied at the *address and intersection* level to create homogenous areas to assign subsequent patrols (Deryol et al., 2016). Although raster based approaches to predictive policing are popular (Caplan & Kennedy, 2011), these raster grids do not conform to the actual micro places and the street network that heavily influence crime (Groff, 2014; Rosser et al., 2017). While prior work has used the creation of the raster grid orientation and size of the grid cells as a hyperparameter when tuning machine learning models (Flaxman et al., 2019; Mohler & Porter, 2018), one could similarly use the DBSCAN parameters of epsilon and minimum weight in the same capacity to avoid relying in grid cells entirely. Such results are likely to result in more natural, contiguous hot spot areas to target. Such contiguous areas would not only require fewer resources to target (relative to separate hot spots not in the same area), but also would likely result in better police adherence when targeting those hot spots boundaries (Sorg et al., 2017).

-Pg 13-14: Not being as familiar with Dallas, should sub-area analyses be considered based on the road design – grid in the inner city and moving towards stem and leaf as the city moves towards suburbs – changing the average street length and also the likelihood of certain crimes over others based on ambient population differences

Response: I do not understand this suggestion. Are you saying I should change the nature of the DBSCAN algorithm to allow a larger epsilon in different areas of the city? Or are you saying my results are

misleading since I average over the entire areas? Creating my own algorithm is a bit much for Police Quarterly to be frank.

Results:

-Pg 15-16: Any knowledge in/how TAAG were generated across the city and what data were used?

Response: Ditto from earlier, I do not have any special insight.

-Pg 16-17: Do the author(s) have access to time spent on calls? Would be an interesting addition from a practitioner standpoint to understand how much officer time is spent on each street along with the crime harm PAI

Response: While I do have access to this, this would require a pretty substantial addition to the analysis. (It isn't as simple as insert a line of code and out pops the results.) This is beyond what is reasonable to ask for a revision.

-Pg 17: "This particular how spot contains..." Could briefly tie these to known crime generators/attractors with the possibility of being risky facilities – leading to crime clusters there

Response: I have amended the sentence to now say this:

This particular hot spot contains a Family Dollar store, a gas station, an elementary school, and several apartment complexes, each likely an individual high risk crime generator (Eck et al., 2007), that all in toto contribute various crimes to this hotspot.

-Pg 18: Could the differing findings from Haberman also be due to site differences? If we treat each study area as unique, some of the general findings need to be tested across multiple sites – could be an example of limited findings based on city context

Response: I have added in this sentence to mention that non-generalizability of the findings may also explain the differences.

Although the difference in results may also be the case that Dallas has substantively different crime patterns than Philadelphia.

Discussion:

The points brought up fit the paper as a whole, especially with the policy implications. Personally, I would like to see language geared towards the feasibility of analysts to conduct similar analyses within an agency – a translational part aimed at the analytical nature of the current study

Response: I have added this paragraph to end the discussion.

The feasibility of such an approach should be within the capabilities of Dallas PD to implement. This analysis comes with a set of replication materials using open source data and open source code, thus Dallas PD (or any other interested police department) can replicate the analysis using more current data for free if they so wish. Given the consistency of other papers identifying hot spots of crime, it is likely the case such work can generate areas of hot spots which show promise of generating significant returns on investment for police departments. Hot spots weighted by cost of crime estimates can then be used as an upfront tool to justify either

new investments in PDs to allocate resources to hot spots or to shift current resources to hot spots. This provides a direct cost-benefit calculation to justify to police departments to shift resources from reactive to proactive policies (Nagin et al., 2015).

Reviewer: 3

This paper introduced/applied a new clustering technique, DBSCAN, for identifying hot spots, that allows the analyst to weight crimes, in this case using cost of crime estimates, and demonstrates its superiority for predicting crime compared to Dallas PD official hot spots. This paper is an interesting idea for all the contributions the authors list in the paper and I support publication if the paper is revised.

Did the authors consider comparing DBSCAN to other hot spotting techniques too? For example, Drawve (2016) A Metric Comparison of Predictive Hot Spot Techniques and RTM in JQ might be a useful framework to think about DBSCAN in a wider context. Overall, the Dallas PD hot spots do not seem that useful for comparison purposes as the authors already knew they were quite large in the first place... I think contextualizing DBSCAN within a wider hot spots framework might make a bigger contribution. I wouldn't go overboard but adding some of the "usual suspect" technique types might be enough.

Response: We did not here. We have added this lack of comparison to other methods as a limitation:

Thus while we cannot say that our DBSCAN analysis is dispositive that it is better than these other approaches, we believe it is likely the case one can use many of these different approaches to improve upon the current Dallas TAAGs, even simply just counting up street segments with the highest crimes is likely much more accurate (MacBeth & Ariel, 2017; Wheeler & Steenbeek, 2020).

Note this is a common critique of policing research that I think is actively harmful – you do an analysis that shows a PD is doing something very inefficient/ineffective, and other criminologists say 'Oh we all know that is bad, no need to write a paper on it'. It is important to address such ineffective practices though, even if we know going in that they are bad. Quantifying how much better an alternative is I believe is sufficient to justify the publication. We already admit in the paper that many different ways counting up crime is likely to be better suited than the status quo. Adding in additional analysis of different ways to be more effective does not change the fact that this way we have described here is quite a deal better than the current status quo.

We have another publication that does compare various hot spot metrics with similar Dallas data (although not weighted by crime harm), Wheeler & Steenbeek (2020). We find that random forests and simply counting crime are far superior to either RTM or kernel density estimates, so I doubt our analysis here would be greatly improved by using any of the 'usual suspects'. We have provided the data and code to replicate, so if others want to do additional work the data is there for the authors to take that step.

And we discuss already in the manuscript in the discussion how one might marry regression based hot spot forecasts with DBSCAN (which are simply crime in and crime out). That paragraph is copied below.

While this work focuses on long term crime forecasting, the application of using DBSCAN can potentially be extended to short term crime forecasting (Flaxman et al., 2019;

Garnier, Caplan, & Kennedy, 2018; Lee et al., 2019; Mohler et al. 2015; Ratcliffe et al., 2020). One approach that we believe may be fruitful is to use DBSCAN on the output of a predictive policing application applied at the *address and intersection* level to create homogenous areas to assign subsequent patrols (Deryol et al., 2016). Although raster based approaches to predictive policing are popular (Caplan & Kennedy, 2011), these raster grids do not conform to the actual micro places and the street network that heavily influence crime (Groff, 2014; Rosser et al., 2017). While prior work has used the creation of the raster grid orientation and size of the grid cells as a hyperparameter when tuning machine learning models (Flaxman et al., 2019; Mohler & Porter, 2018), one could similarly use the DBSCAN parameters of epsilon and minimum weight in the same capacity to avoid relying in grid cells entirely. Such results are likely to result in more natural, contiguous hot spot areas to target. Such contiguous areas would not only require fewer resources to target (relative to separate hot spots not in the same area), but also would likely result in better police adherence when targeting those hot spots boundaries (Sorg et al., 2017).

Did the authors do any supplementary analysis using DBSCAN without weights? In other words, how do “raw” hot spots compare to the weighted hot spots? Same number? Same general locations? Same PAIs? In other words, while I agree weighting by cost of crime has its own substantive import, does it make much of a difference at the hot spot identification phase or could it have the same import if applied post-identification to hot spots identified using traditional methods?

Response: We did not do DBSCAN without weights, but we do demonstrate that DBSCAN with weights generates near equivalent non-weighted PAI statistics (see the crime specific rows in Table 2). So even if one is not happy with the weights, this still suggests the technique does quite well in identifying hot spots for just the crime counts. One could also filter those particular results to only choose areas above a particular violence threshold if they so wanted.

What would the authors say to police officials who care more about reducing raw crime numbers? What about folks who raise concerns about qualitative differences in crime types? We all know police chiefs are most accountable to homicides/shootings due to the cost of human life. Should these points be considered more formally in the discussion?

Response: Similar to the prior response, this probably presents a false dichotomy. The high crime cost hot spots are similarly areas of high violence according to the same PAI metric. We have added in this paragraph in the discussion to make this point:

It may also be that a police department wishes to focus on a particular set of crimes for a hot spots policing strategy, e.g. just focus on gun violence. Given our identified hot spots here tend to have a variety of different crimes, it may still make sense to generate general cost of crime hot spots, and then filter for those areas that contain certain levels of specific crime types, as it is likely the case an intervention focused on one crime type is likely to have positive spillovers to reducing other crime types.

The authors might consider providing more descriptives about raw crime counts for the clusters in the tables. Something like Table 2/4 in the Haberman (2017) paper cited. This would provide some context for readers.

Response: The equivalent of Table 4 are available in the supplemental replication materials as a CSV file, see the ClusterStats.csv file. I personally find it not that informative, it is just a table of crime counts for 100+ areas. I leave it up to the editor if he wishes to include this in the manuscript, seems to detailed in the weeds in my opinion. Tables 2 & 3 in Haberman are not relevant here due to the differences in the nature of the analysis.

The authors might consider providing a more formal treatment of the DBSCAN statistic for readers who have not encountered it before. After all, introducing it to CJ/policing is a contribution of the paper. Clear limitations of the technique might be helpful for readers encountering the technique for the first time.

Response: Here is the updated section on DBSCAN, in particular the second paragraph establishes the limitations:

For a simplified example of how DBSCAN works, imagine you have two locations, A and B, that are 300 feet apart, and each have a total of \$200,000 in crime cost. Even though each location by itself does not meet the \$400,000 threshold, combined they do. The epsilon parameter is what defines the threshold as to whether two points can be combined to consider their joint weight. Points A and B then form a *core cluster*. Points additionally within 400 feet from either A or B will additionally be considered inside the cluster, but unless all points within a particular radius of 400 feet exceed the \$400,000 in crime cost, those points will not be considered a core point for the cluster.

Thus the epsilon distance parameter and the minimum weight parameter are two details that need to be chosen by the analyst, they are not automatically chosen in an algorithmic way. This is both a strength and a weakness – there are likely no universal solutions for what those values should be to produce the ‘best’ hot spots, but that choice also allows the analyst to adjust the parameters for their own particular circumstance. So cities that are more spread out an analyst may choose a larger distance parameter, or for hot spot interventions that are more costly the analyst may make the threshold for hot spot identification larger as well. Ultimately *all* hot spot creation procedures involve some arbitrary decisions, e.g. choosing a bandwidth for a kernel density map, and the necessary decisions to use DBSCAN is no more onerous than other common hot spot techniques.

While readers can go directly to the Hunt et al. (2019) piece, it may be a good idea to more explicitly describe how the cost of crime estimates were computed given their importance for this piece. Additionally, the authors should consider how much of the costs would be direct benefits to police departments vs. society in general. Does where those costs come back to make a difference to the policy discussion? In other words, how would knowing the costs of crime impact police departments’ own bottom lines?

Response: I have added in this section to further detail the Hunt methodology.

To briefly describe Hunt et al.'s (2019) methodology in more detail, it specifically focuses on the labor costs of *responding* to crime. Criminologists often distinguish between reactive and proactive policing (Nagin et al., 2015), and Hunt's estimates are relevant for the reactive category. The reactive police response to crime can be broken down in-between various categories; Hunt et al. (2019) list different categories administrative, arrest, crime scene, court time, investigative, and en-route/waiting. They generate state level estimates for costs of responding to crime based on various data sources, including Bureau of Justice Statistics resources for justice expenditures and time spent responding to different crimes (broken down by differing roles, such as patrol officer vs detective), and state crime incident totals from the Uniform Crime Reporting program. They further estimate breakdowns between urban and rural agencies using individual level time diary estimates from studies specific jurisdictions, and conduct monte-carlo simulation to generate distributions around point estimates, although in this study we only use the point estimates for Part 1 crimes.

We already have a section specifically outlining the different ways to calculate harm in the discussion, and how these estimates likely *underestimate* costs of crime if one is further considering general costs to society. That section is copied below:

One strong limitation of the current analysis is that the definition of the cost of crime is not well agreed upon. While here we use the work of Hunt et al. (2019) that is mostly based on costs associated with police response to particular crimes, there is a wide array of ways to calculate the cost of crime (Domínguez & Raphael, 2015). For example, a murder in this analysis is valued at \$124 thousand dollars, whereas the value of a statistical life is often pegged at over one million dollars (Domínguez & Raphael, 2015). This difference has to do with who bears the cost of the murder – for police it is mostly in terms of investigative time, whereas the cost is much more severe for the victims family (and infinite for the victim themselves). It may also be that crime prevented for individuals accumulate over a lifetime (Cohen & Piquero, 2009), with failing to prevent crime resulting in long term negative externalities for various individuals. Finally, this does not consider additional considerations on allocating police resources, such as equity or preventing disproportionate minority contact with a hot spots policing approach (Wheeler, 2019). So while these cost of crimes estimates are relevant to police departments themselves, they may not be reasonable to the greater public when constructing a hot spots policing strategy.

Additionally, this work relying on public data does not include rapes, and we also do not include Part 2 crimes, which are often considered public disorder crimes and additionally contribute to quality of life (Boggess & Maskaly, 2014; Chappell, Monk-Turner, & Payne, 2010; Ratcliffe, 2015). This is not a limitation for a police department applying such estimates internally, where they have access to all of the reported crime data, but even then non-reporting will cause such estimated cost of crime hot spots to be an *underestimate* of the cost of crime at particular places. Ultimately how to appropriately value crime harm is as much an ethical question as it is an empirical. This work shows one way to do so, using costs directly relevant to police departments, but this is not to say other ways, such as via sentencing severity or public surveys are inherently inferior. But it is likely the case that many of these different ranking schemes will produce similar rankings in crime (e.g. all would likely identify homicide as a higher harm than theft), so we believe using *any* crime harm ranking scheme is likely better than the defacto standard of treating such crimes equally when creating hot spots.

So we believe that both of these biases likely underestimate the cost of crime that is occurring the hot spots we identify here. Given that the ultimate goal of identifying such hot spots is to prevent crime, these estimates may then be gross underestimates of the potential return on investment if Dallas PD targets crime in those identified areas. But that is also conditional on a hot spots policing strategy to be effective (Mitchell, 2019) – just placing hot spots on the map and not doing anything with that information will ultimately not reduce crime at those locations.

Can the authors describe the heterogeneity of crime types more formally? Perhaps a heterogeneity index would be informative here? Or maybe descriptive for the number of crime types included in a hot spot across all hot spots?

Response: I calculated the entropy statistic to address this. Below is the updated part of the manuscript (including the footnote for the calculation):

To quantify this, we have estimated the Shannon entropy statistic for each of our DBSCAN clusters (Lee & Eck, 2019).¹ For this statistic, if a hot spot only contained one crime type, the entropy would be 0. Higher values signify more entropy, so it is harder to predict what type of crime may fall within that hot spot. The average entropy for each area is 1.38, with a maximum entropy given six potential crime types under examination here is 1.79. Thus areas are much closer to max entropy than they are encompassing a predictable set of crime types.

The interactive map is nice, but the paper will likely outlast the authors' website. I would encourage the authors to make sure the static maps in the paper can stand along and perhaps improve them a bit with some additional cartographic details (e.g., scale bar, north arrow, etc.).

Response: It is on GitHub, which is owned by Microsoft, not my personal website. I will additionally deposit the replication materials as supplemental files to Police Quarterly. I have included in the screenshot a scale bar, and north is up as usual (it is atypical to include a north arrow in sloppy webtile maps).

Figure 2 is really nice.

Response: Thank you.

¹ The entropy statistic can be calculated as $E = -\sum_c p_c \cdot \log(p_c)$, where p_c is the proportion of crime c out of all crimes in that hot spot area. So for an example, a hot spot location of 5 robberies, 3 murders, and 2 motor vehicle thefts would have an entropy statistic of $E = -[\frac{5}{10} \cdot \log(\frac{5}{10}) + \frac{3}{10} \cdot \log(\frac{3}{10}) + \frac{2}{10} \cdot \log(\frac{2}{10})] \approx 1.02$. Crimes with zero values do not enter into the calculation.