Crime Mapping & Analysis News

A POLICE FOUNDATION PUBLICATION

Issue 2: Spring 2015

Mapping Changes in Neighborhood Crime Trends Post-Disaster

Crime Forecasting on a Shoestring Budget

Data and Research for Increased Safety and Fairness

Foreclosures, Domestic Disturbances, and Policy Implications

Using Risk Terrain Modeling Technique to Identify Places with the Greatest Risk for Violent Crimes in New Haven

www.crimemapping.info www.policefoundation.org
Crime Mapping & Analysis News
A POLICE FOUNDATION PUBLICATION

Issue 2: Spring 2015

Jim Bueermann, President
Shefali Tripathi, PhD, Editor-in-Chief
Karen L. Amendola, PhD, Editor
Laura A. Wyckoff, PhD, Advisor

Jim Specht, Communications Manager
Mary DeStefano, Creative Communications Manager

*Front cover image art provided by Shefali Tripathi, Gainesville Police Department, FL.

About the Police Foundation

The Police Foundation is the only nationally-known, non-profit, non-partisan, and non-membership-driven organization dedicated to improving America’s most noble profession – policing. The Police Foundation has been on the cutting edge of police innovation for 45 years since it was established by the Ford Foundation as a result of the President’s Commission on the Challenge of Crime in a Free Society.

The Police Foundation, relying on its in-house staff and a team of executive, policing, and research fellows, provides actionable technical assistance and conducts innovative research to accomplish its mission of improving policing through science and innovation.

Our Partners:
Welcome to the second issue of the Crime Mapping & Analysis News, and the first of 2015. I am happy that the first issue was well received by our readers.

This issue brings a broad array of articles focusing on crime mapping and analysis practice and research. The articles included in this issue are:

- Mapping Changes in Neighborhood Crime Trend Post-Disaster
- Crime Forecasting
- Importance of Data and Research in Criminal Justice
- Foreclosures, Domestic Disturbances, and Policy Implications
- Risk Terrain Modeling

I am excited to bring these new topics to our readers and I believe that you will find them informative.

I sincerely appreciate the continued support of our contributors and readers alike. I would also like to thank the editorial team at the Police Foundation for their help and support in bringing out this exciting issue. If you have feedback about this issue or would like to submit an article, please feel free to contact us at editors@policefoundation.org.

Sincerely,
Shefali Tripathi
Editor-in-Chief
In recent years the world has experienced a number of extreme disaster events. The 2004 Indian Ocean Tsunami devastated countries from East Africa to Thailand, and in the United States, Hurricane Katrina in 2005 had far reaching and long lasting consequences for New Orleans. In 2011, Christchurch, NZ was severely damaged by one of the largest earthquakes recorded in the country. This event was closely followed by a tsunami that ravaged large areas of Japan. Natural disasters resulting from rapid onset hazards such as floods, bushfires, and storms incur wide-ranging impacts on individuals and communities. There is some consensus that disaster events are increasing in frequency and severity (Gencer, 2013; World Bank, 2010). Further evidence suggests that natural disaster costs are increasing and will continue to rise in the future, as a function of population growth, an aging housing stock, and growing concentration of assets in disaster prone-areas (Australian Government Productivity Commission, 2014; World Bank, 2010). These costs are direct (e.g. damage to public and private property and infrastructure) and indirect (e.g. flow on effects as the community responds to the disaster). A potential indirect cost of disaster is increased crime.

Although some studies suggest disaster increases community capacity for social regulation by bringing residents together, through their shared experience of trauma (Drabek & McEntire, 2003; Quarantelli, 2005), others find disasters produce anomic conditions that encourage people to panic. This can then hinder normative behaviors that are necessary for informal social regulation, leading to an increase in crime (Erikson, 1994). Further disasters alter the routine activities of residents and can increase the number of properties left unguarded, which in turn increases opportunities for crime (Leitner and Guo, 2013). To mitigate the negative effects of a disaster event it is imperative to understand what makes a neighborhood vulnerable to negative outcomes, including crime increases in the post disaster context.

In January 2011 Brisbane, the state capital city of Queensland, Australia, experienced an extreme flood event. It impacted over 175,000 people, and resulted in $7.5 billion dollars worth of damage, making it the most costly natural disaster in Australia’s history. Thousands of homes were damaged or completely destroyed, and the composition of neighborhoods changed as a result of the flood. This study investigates whether these changes to the neighborhood structure influenced crime trends across 390 communities in Brisbane, following the flood event. It examines the association between post-flood changes in property crime and neighborhood structural characteristics, and changes in those characteristics and crime in nearby neighborhoods. In so doing it addresses three questions: what pre-disaster neighborhood characteristics indicate vulnerability to post-disaster crime increases?; how do neighboring

Renee Zahnow is a PhD candidate and Research Assistant at the University of Queensland. She is currently in the final stages of her PhD. Her research interests focus on the physical and social contexts in which crime occur. She is specifically interested in how neighborhoods change over time and how that change is related to crime and disorder.
communities influence post disaster changes in property trends; and, do disaster related changes in neighborhood characteristics influence property crime trends? This research advances earlier work on disaster and crime by examining changes in crime trends post-flood and by incorporating spatial effects.

Methods

Geographical Information System (GIS) was used to spatially integrate Australian Bureau of Statistics census data and property crime count data with maps depicting flood extents. ARIMA times series analysis was employed to model neighborhood crime using pre-flood monthly crime count data (1996-2010) and to forecast expected post-flood crime trend. The forecast trend was then compared to actual crime counts to establish whether or not post-flood property crime trend was significantly different than forecast. A neighborhood was found to experience post-flood property crime that diverged from pre-flood trend if post-flood property crime fell outside of the 95 percent confidence interval for the forecast trend (for example see Figure 1). After conducting ARIMA times series analysis for all Brisbane neighborhoods each was assigned to one of three categories: post-flood crime not significantly different than forecast; post-flood crime lower than forecast; post-flood crime higher than forecast. Multi- nomial logistic regression was employed to assess structural characteristics associated with an increased likelihood of deviation from forecast crime post-flood.

Results

The results of the ARIMA time series analyses indicated that in most neighborhoods (70%) property crime trend did not deviate from forecast post-flood (Figure 2). Both flooded and non-flooded neighborhoods were among those that experienced deviation from forecast property crime trend post-flood. While there was no significant spatial clustering in deviation from forecast crime trend (Moran’s I= 0.055), pre-flood spatial context was a predictor of post-flood deviation from crime trend. Specifically, neighborhoods that were surrounded by low crime areas pre-flood were more likely to experience post-flood crime that was lower than forecast. Other neighborhood characteristics associated with greater likelihood of lower than forecast crime post-flood were higher population density and lower residential mobility. Higher residential mobility, percentage households renting and percentage single parent households were associated with greater likelihood of a
neighborhood experiencing higher than forecast crime post-flood.

This research uses spatially integrated data to examine the effects of disaster on crime and to identify neighborhood characteristics associated with greater vulnerability to post-disaster crime increases. This study found that residential mobility, population density, percentage of households renting and percentage single parent households predicted greater likelihood of post-flood deviation from forecast crime trend. It also found that pre-flood crime in the neighborhood of interest and in surrounding neighborhoods was associated with likelihood of deviation from forecast crime trend in the disaster context. These results have a number of implications for future research and policy regarding disaster preparedness and response. Pre-disaster identification of places most vulnerable to crime increases can assist with resource allocation and ensure preparedness and prevention actions are focused on these neighborhoods. The findings suggest that following a disaster event it may be necessary to increase police patrols in particular types of places to prevent unexpected increases in property crime. It may also be necessary to educate residents in vulnerable neighborhoods on the increased risk of property crime in the post-disaster period. There is potential for this analysis to be expanded to incorporate indicators of density of crime attractors and generators to assess whether the presence of particular physical characteristics make neighborhoods more vulnerable to experiencing changes in crime trend in the event of a disaster. A limitation with existing research on the effect of disaster on crime trends is that it does not control for spatial variations in key structural and social characteristics that are known to influence patterns of crime in times of disaster quiescence. This research begins to redress this gap in the existing scholarship.
Confidence interval for the forecast trend (for example see Figure 1). After conducting ARIMA times series analysis for all Brisbane neighborhoods each was assigned to one of three categories: post-flood crime not significantly different than forecast; post-flood crime lower than forecast; post-flood crime higher than forecast. Multi-nomial logistic regression was employed to assess structural characteristics associated with an increased likelihood of deviation from forecast crime post-flood. Specifically, neighborhoods that were surrounded by low crime areas pre-flood were more likely to experience post-flood crime that was lower than forecast. Other neighborhood characteristics associated with greater likelihood of lower than forecast crime post-flood were higher population density and lower residential mobility. Higher residential mobility, percentage households renting and percentage single parent households were associated with greater likelihood of a deviation from crime trend. Specifically, neighborhoods that were surrounded by low crime areas pre-flood were more likely to experience post-flood crime that was lower than forecast. Other neighborhood characteristics associated with greater likelihood of lower than forecast crime post-flood were higher population density and lower residential mobility. Higher residential mobility, percentage households renting and percentage single parent households were associated with greater likelihood of a deviation from crime trend.

References


This research uses spatially integrated data to examine the effects of disaster on crime and to identify neighborhood characteristics associated with greater vulnerability to post-disaster crime increases. This study found that residential mobility, population density, percentage of households renting and percentage single parent households predicted greater likelihood of post-flood deviation from forecast crime trend. It also found that pre-flood crime in the neighborhood of interest and in surrounding neighborhoods was associated with likelihood of deviation from forecast crime trend in the disaster context. These results have a number of implications for future research and policy regarding disaster preparedness and response. Pre-disaster identification of places most vulnerable to crime increases can assist with resource allocation and ensure preparedness and prevention actions are focused on these neighborhoods. The findings suggest that following a disaster event it may be necessary to increase police patrols in particular types of places to prevent unexpected increases in property crime. It may also be necessary to educate residents in vulnerable neighborhoods on the increased risk of property crime in the post-disaster period. There is potential for this analysis to be expanded to incorporate indicators of density of crime attractors and generators to assess whether the presence of...

By Andreas M. Olligschlaeger

Unless you’ve been living under a rock for the past three or four years you’ll have heard of predictive policing, the latest of many buzzwords to emerge in the field. Media hype has been nothing short of spectacular, with many vendors – new and old – jumping into the fray and offering software products that promise to revolutionize policing as we know it. Naturally, all of this comes at a cost: not only are some of the products extremely expensive, but they have created sometimes unrealistic expectations of crime analysts. In a few rare instances city and police managers have been convinced that software can somehow reduce the need for human analysts or even replace them entirely.

Unlike the term predictive policing, crime forecasting has been around for a long time (see for example Olligschlaeger, 1998). What is new, however, is that efforts are being made to incorporate crime forecasting techniques – some new, some old – into daily police operations. This, in essence, is what predictive policing is, although some of the techniques used overlap with crime analysis concepts that have been in use for decades.

Unfortunately the media hype has led many to believe that predictive policing is a product that can be bought off the shelf. However, as Perry et al. (2013) and others have pointed out, predictive policing is not a software package or a statistical technique. Rather, it is a complex process, of which software and statistical methodology are just one part - albeit an important one. The process of using software and analysis to produce forecasts of criminal activity is what has traditionally been known as crime forecasting.

There are many crime forecasting methodologies, depending on what it is a crime analyst wants to forecast. They range from the very simple to the highly complex. It is important to note that highly complex models – those typically sold in commercial packages - often only provide an incremental improvement in terms of accuracy and utility over some of the simpler or intermediate methods. This means that a well-trained analyst can in many instances produce results almost as good as a sophisticated commercial software package costing a lot of money with tools that are either already in his or her toolbox, can be purchased for a relatively small amount of money or are available open source. For increasingly cash strapped police agencies this brings up an important question: does it make sense to spend large sums of money on commercial software or would it make more sense to provide training to crime analysts in order to teach them to build their own forecasting models? Some police agencies, especially small to medium sized ones might benefit more from the latter and realize a greater return on investment in the long run.

This last point brings me to the purpose of this article: how to put together a basic, yet powerful and robust crime forecasting toolkit on a shoestring budget or even no budget at all. I begin by discussing some of the requirements for forecasting (primarily data and software) and provide a very brief overview of some of the most commonly used techniques for one type of crime forecasting: one step ahead forecasts by a real unit. This is followed by suggestions for the crime forecasting toolkit and some examples of how that toolkit might be used in practice.
Crime Forecasting Requirements, Data Issues, and Methods

For simple to intermediate crime forecasting methods — the primary focus of this article — most analysts already have many of the software packages they will need. Nevertheless, it is worth briefly discussing individual components because the sophistication of each will have a bearing on what it is you will be able to do.

The first component is obviously a geographic information system (GIS). While strictly not required to produce actual crime forecasts, it is useful for displaying results as well as collecting data. For example, a GIS could be used to generate frequency counts of geocoded crime incidents by census tract, census block or grid cell. Many GIS systems also have built in statistical capabilities, such as spatial regression and data manipulation routines. The more sophisticated the GIS, the more you can do with your data.

A staple of every crime analysis unit is a SQL compatible database. Databases are essential for gathering, storing and manipulating data that is often retrieved or exported from a variety of other database, such as an RMS, CAD system or others. Naturally, there are a variety of ways in which data can be imported into a database, but depending on the need another good tool to have in the crime forecasting toolbox (or in the crime analyst toolbox in general, for that matter) in conjunction with a SQL database is a data migration tool. Data migration tools are very useful for setting up and conducting automated data transfers/exports on a regular basis from multiple databases. The advantage of such tools is that they can connect to multiple platforms, even if they are in the cloud, and allow the user to manipulate data before they are deposited into the crime analysis database. For example, date/time formats can be changed, fields parsed etc., meaning that many tasks a crime analyst normally has to perform manually can be automated. So they are also a significant time saver. But even if a crime analyst does not have direct access to an RMS or CAD database and depends on, say, regular XML export files, a data migration tool can be set up to automatically scan a local or network directory for a new export file and, if it finds one, process it.

Finally, no crime analysis unit should be without a good statistical software package. Most GIS systems and standard productivity software such as Microsoft Excel have some statistical features built in to them, but in order to progress beyond the most simple of crime forecasting techniques something more powerful is needed. Commercial statistics packages can be very expensive but fortunately there are some excellent open source alternatives.

One of the issues unique to crime forecasting as compared to forecasting in other disciplines is that crime is a rare event, yet successfully building a crime forecasting model can require hundreds or thousands of data points. The more data you have and the higher quality it is, the more sophisticated a forecasting model you can build. So the first and most important step is to decide what types of data to use, how to get it, and how to incorporate it into a single dataset.

A further issue that is related to how many data points you have is choosing the size of area by which you want to forecast as well as the time frame. While many analysts will have no choice but to use census tracts or blocks, or even police beats, those who decide to use grid cells or some other variable area size will have to experiment. If you choose too large an area you will have plenty of crimes but the tactical utility of even accurate forecasts diminishes. Pick too small of an area and most areas will have zero or one crime counts, meaning most forecasting methods will treat areas of high crime as outliers. Naturally, what
The primary goal of crime forecasting is not just to create forecasts that are accurate, but more importantly forecasts that tell police officers what they don’t already know. For example, accurately predicting that historical crime hotspots will remain hotspots in the next time period doesn’t provide much tactical insight. Neither does accurately predicting that most census blocks will have zero homicides next month. What is really needed are models that can predict new hotspots before they emerge, although almost equally as important is being able to forecast decreases in crime – after all, both have implications for the criminal justice system, depending on what the purpose of the forecast is (see Perry et al., 2013, for a more complete overview). One of the most common types of forecast produced by crime analysts is the one step ahead forecast by area. In other words, we want to forecast either levels of crime or changes in crime in the next time period (usually week or month) by area (census tract, census block, or grid cell). There are many methods that can accomplish this, varying from simple (such as simple exponential smoothing and Holt Smoothing) to intermediate (multiple regression) to highly complex (artificial neural networks and genetic algorithms) (see for example Gorr and Olligschlaeger, 2002).

There are two basic types of forecasting method used in one step ahead forecasts: univariate and multivariate. Univariate methods are simpler and require fewer data points. Typically they require only a time series of crime counts by area. Multivariate methods, on the other hand, incorporate data from a variety of sources such as RMS, CAD, public property records, and others. Those methods, such as regression, for example, are usually more accurate and better at identifying new areas of crime because they incorporate variables that can contribute to increases or decreases in crime. Such variables are also known as leading indicators (Cohen et al., 2007).

The primary goal of crime forecasting is not just to constitute the best size areal unit will depend on the crime type as well as jurisdiction.

The choice of temporal unit also affects how many crimes you will have per unit for much the same reasons as areal units. Typically forecasts are either weekly or monthly, with monthly the most commonly used.

There are many different types and categories of forecasting or predictive models used by police and the criminal justice system, depending on what the purpose of the forecast is (see Perry et al., 2013, for a more complete overview). One of the most common types of forecast produced by crime analysts is the one step ahead forecast by area. In other words, we want to forecast either levels of crime or changes in crime in the next time period (usually week or month) by area (census tract, census block, or grid cell). There are many methods that can accomplish this, varying from simple (such as simple exponential smoothing and Holt Smoothing) to intermediate (multiple regression) to highly complex (artificial neural networks and genetic algorithms) (see for example Gorr and Olligschlaeger, 2002).

While discussing the various methods in detail is beyond the scope of this article, there are a number of general tips that apply to most multivariate methods:

- Try a number of different models to see which works best.
- Use a holdout sample (a portion of the data that was not used to calibrate the forecasting model) to compare how well the model generalizes. If the model performs as well or almost as well on the holdout sample as it does on the data used to estimate the model then a model is said to generalize well. Otherwise your model may be over fitting the data.
- Thoroughly analyze and explore your data before you build a forecasting model. Look for leading indicators that are correlated with what it is you are trying to forecast and avoid throwing as many variables into the model as possible.
Overall model fit is not as important as being able to identify emerging or declining hot spots.

If you have evidence of geographic displacement of crime, think about using spatially and temporally lagged variables, i.e. variables from neighboring areal units from the previous time period. The same is true if events in one area influence those in neighboring ones.

Compare the results of your model to simple methods, such as the random walk, for example. The random walk assumes that the crime count in the next time period is the same as in the previous one. The random walk can be surprisingly hard to beat.

Forecasting Toolkit

Now that we’ve looked at what we need to be able to do with our data and what types of software we need, it’s time to put together the actual crime forecasting toolkit. The good news is that just about everything is available from open sources. While some open source resources require at least a basic knowledge of scripting and programming languages (usually Java), there are a plethora of how-to manuals available, many of which are free. A big advantage of open source products is that there are usually numerous forums, support, and user groups associated with each product. This means that if you are looking to find a script or program to do something, the chances are pretty good that someone else has either already done it or has at least done something similar to what you’re looking for and all you have to do is search the web for it. But most importantly, the term “open source” does not necessarily mean cheap (as in you get what you pay for), inferior, or substandard products. Quite to the contrary, many of the best known websites including Twitter, Facebook, Yahoo, and Wikipedia are entirely or in part developed and deployed using open source products. For those that are concerned about security and vulnerabilities associated with using open source products, rest assured. They are as good if not better than most commercial products.

While there are numerous open source GIS packages, none are likely as powerful as the GIS that most crime analysts already have, which at the time of writing is usually ArcGIS. However, open source GIS packages are available that approximate at least some of the capabilities of commercial GIS packages, as well as web map servers and other geospatial tools. As most other open source products many of these will run on multiple platforms, including Windows, Linux and MAC operating systems. For a complete listing of open source GIS software, their capabilities as well as links to where they can be downloaded see http://en.wikipedia.org/wiki/List_of_geographic_information_systems_software.

Most crime analysis units have access to at least one SQL compatible database, usually Microsoft Access. While Access – especially the latest versions – will suit most analysts just fine, there are a number of very sophisticated and high powered open source database available. Of those, MySQL (http://dev.mysql.com/) and PostgreSQL (http://www.postgresql.org/) are the most popular and sophisticated, easily rivaling some of the most prominent commercial products, including enterprise products. For those of you using ArcGIS there are connectors available to both databases. Using a more sophisticated open source database makes sense for those analysts that consistently deal with large amounts of data, need to create a data warehouse, or need to share data with other units or analysts.

Data migration tools are extremely useful for those analysts that regularly receive data in the form of direct SQL queries or export files from one
or more databases such as CAD, RMS and others. Crime analysts in general spend too much time focusing on data collection and integration and not enough time doing analysis. With a little bit of work (and perhaps some help from the IT department) it is easy to set up routine data exports/imports from and to other databases. Many migration tools have simple drag and drop interfaces where all you need to do is fill in parameters such as export file locations, database credentials, SQL commands, etc. Those analysts with even basic programming skills can add as much sophistication as desired, like data transformations, SQL commands to automatically increment crime counts by areal unit and time period based on new CAD and/or RMS incidents coming in, etc.

The two most commonly used data migration tools are Mule (www.mulesoft.com) and Apache Camel (camel.apache.org). Both can connect to all common commercial and open source databases (even in the cloud) and run on any popular operating system. Mule and Camel also both work with the two most widely used open source integrated development environments (IDE’s): Eclipse and NetBeans. For example, you can download a version of Mule that is already bundled with Eclipse and ready to use as soon as you have installed it.

The final piece of software in our crime forecasting toolkit is a decent statistical package. There are numerous general purpose open source statistical packages available, as well as some government funded ones that are free. Of the latter the one that really stands out is CrimeStat (http://nij.gov/topics/technology/maps/Pages/crimestat.aspx). CrimeStat was developed specifically for crime analysts and contains many spatial statistical routines useful for crime forecasting. While crime analysts have been using CrimeStat for years, it is worth mentioning because the latest version – CrimeStat 4 – includes two types of univariate forecasting, simple exponential smoothing and Holt smoothing as well as a Trigg signal detector that can identify areas that exhibit signs of unusual activity.

As far as general open source statistical packages are concerned, there is a plethora available (see http://en.wikipedia.org/wiki/List_of_statistical_packages for a detailed list). However, one in particular – R (http://www.r-project.org/) – easily stands out. R not only has highly complex statistical routines that can be used for crime forecasting and data exploration, but it can also produce high quality graphics. In addition to user groups, tutorials and forums a number of publications can help the budding crime forecaster get started (see for example Lander (2014) and Zumel and Mount (2014)). An excellent open source tutorial by Rob Hyndman and George Athanasopoulos of Monash University in Australia on using R for forecasting that explores some of the simple, intermediate and even advanced methods such as artificial neural networks can be found at https://www.otexts.org/fpp.

Finally, RStudio (http://www.rstudio.com/) is a powerful integrated development environment for R that will work on all major operating systems and includes tools that allow you to create interactive reports and visualizations for the web. Perfect for disseminating forecasts and other crime analysis products.

**Crime Forecasting Examples**

There are six steps involved in producing multivariate one step ahead forecasts by areal unit. They are:

1. Aggregate – obtain data by areal unit and time period. The more data you have, both in terms of the number of variables and the length of the time series, the better.
2. Explore – pick a crime that you want to forecast and analyze the data for other variables that might impact changes in that crime.

3. Explain – use theory combined with data analysis and explanatory model such as regression to identify leading indicators.

4. Forecast – develop forecasting models that use the leading indicators to produce one step ahead forecasts of crime.

5. Feedback – learn from forecast errors and identify possible sources of error.

6. Make changes to forecasting model, if any, and re-estimate for the next time period.

For univariate models steps 2 and 3 can be omitted. Further recommendations for producing crime forecasts are listed in Gorr and Olligschlaeger (2002).

One good way to explore your data is to use simple forecasts, apply the Trigg tracking signal and visualize the results on a map (see Gorr and Olligschlaeger, 2013a and 2013b). The Trigg tracking signal helps to identify areas showing unusual levels of crime (either increases or decreases).

Table 1 is an example of how CrimeStat 4 outputs results from univariate forecasting algorithms, including the Trigg signal strength and any signal trips. These data are then used to produce output such as that shown in Figure 1, which is a map of Trigg signal trips – meaning that shaded areas had unusually high (red) or low (blue) levels of activity – by census tract. Note that there are three areas where red and blue tracts neighbor each other. This is quite possibly in indicator of geographic displacement. One potential reason for geographic displacement of crime is police activity such as arrests for the crime you are trying to forecast, so police activity in a neighboring census tract would be a potential candidate as a

Table 1: Sample Output From CrimeStat Univariate Forecasting Routines (Source Gorr and Olligschlaeger, 2013b)
leading Indicator. Drilling down into data in other areas showing signs of unusual activity could provide further clues.

Another good idea is to create a matrix with all correlations between the dependent variable (the one you want to forecast) and the independent variables. Those variables that are highly correlated with the dependent variable might be good candidates for leading indicators. Table 2 shows correlations between dependent variable LLDRGTOT (log of total drug calls for service by area and time period) and 11 other variables. For example, the table shows that the number of nuisance bars (NBARS) is positively correlated and the log of median household income (LMDH-HINC) is negatively correlated. Once candidates for leading indicators have been identified, explanatory models such as multiple regression can be used to verify that they are indeed a significant contributor changes in the independent variable.

Once you have identified your leading indicators it’s time to do some actual modeling. As mentioned earlier, it is always a good idea to try different models as well as different combinations of variables to see what works best. Table 3 is an example of what a comparison between different models might look like. The table lists a total of eight different models, including the random walk, which assumes that the number of crimes in the next time period is the same as that of the previous time period. In addition to CCF, which is a neural network based model, three types of regression models (Simple, Poisson and Tobit) are estimated with and without spatially lagged averages. Finally, results are compared to a holdout sample in order to determine how well each model generalizes. One thing that stands out is how difficult it is to beat the random walk, at least as far as overall model fit is concerned. Only the neural network model was able to beat it consistently both in the training data set as well as the holdout sample. With one exception all models generalize fairly well.
For univariate models steps 2 and 3 can be omitted. Further recommendations for producing crime forecasts are listed in Gorr and Olligschlaeger (2002).

### Table 3: Comparison of Different Forecasting Models (Source: Olligschlaeger, 1997)

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Data Set</th>
<th>Holdout Sample</th>
<th>Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Regression</td>
<td>0.7795, 45890</td>
<td>0.7644, 22465</td>
<td>-</td>
</tr>
<tr>
<td>Poisson Regression</td>
<td>0.4079, 197447</td>
<td>0.4302, 88787</td>
<td>-</td>
</tr>
<tr>
<td>Tobit Regression</td>
<td>0.7108, 425369</td>
<td>0.6894, 198034</td>
<td>-</td>
</tr>
<tr>
<td>Simple Regression</td>
<td>0.7802, 45725</td>
<td>0.7644, 22449</td>
<td>-</td>
</tr>
<tr>
<td>Poisson Regression</td>
<td>0.3383, 199444</td>
<td>0.3666, 89661</td>
<td>-</td>
</tr>
<tr>
<td>Tobit Regression</td>
<td>0.6894, 45724</td>
<td>0.687, 320908</td>
<td>-</td>
</tr>
<tr>
<td>Random Walk</td>
<td>0.7725, 49988</td>
<td>0.7620, 24165</td>
<td>-</td>
</tr>
<tr>
<td>CCF</td>
<td>0.8188, 37740</td>
<td>0.7800, 20926</td>
<td>4500</td>
</tr>
</tbody>
</table>

*Note: The R-squared reported for CCF, Tobit, and Poisson, as well as the fit of the holdout sample using the regression coefficients is a simple squared correlation coefficient only, and can thus not be directly compared to the adjusted R-squared of the regression results on the training data set or the pseudo R-squared reported by Stata.*

### Table 4 (Source: Olligschlaeger, 1997)

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Data Set</th>
<th>Holdout Sample</th>
<th>Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Regression</td>
<td>1846 (72.36)</td>
<td>1071 (71.29)</td>
<td>-</td>
</tr>
<tr>
<td>Poisson Regression</td>
<td>3145 (123.94)</td>
<td>1734 (121.38)</td>
<td>-</td>
</tr>
<tr>
<td>Tobit Regression</td>
<td>32309 (521.27)</td>
<td>16113 (503.77)</td>
<td>-</td>
</tr>
<tr>
<td>Random Walk</td>
<td>2336 (91.37)</td>
<td>1313 (91.60)</td>
<td>-</td>
</tr>
<tr>
<td>CCF</td>
<td>1411 (65.04)</td>
<td>1005 (67.49)</td>
<td>1005 (67.49)</td>
</tr>
</tbody>
</table>

Note: letters in parentheses indicate a statistically significant difference at the 0.01 level using a Wilcoxon signed rank test.
However, overall model fit doesn’t necessarily tell the whole story. Remember that one of the goals of crime forecasting is to tell police what they don’t already know. One way to do this – there are others - is to look at areas where crime was zero in the previous time period and non-zero in the current one. Table 4 shows a comparison of the same models shown in table 3 for those instances. It is clear that both the simple regression and neural network models beat the random walk and that including spatially lagged averages in regression models improves results, at least for this particular data set. Also worth noting is that while the most complex method – neural networks – performs the best overall it is only marginally better than simple regression. This reinforces the notion that intermediate methods can produce results almost as good as complex ones.

## Conclusion

While this article only briefly touched on some of the many crime forecasting methods available, the reader will hopefully have drawn the conclusion that crime forecasting within the context of predictive policing is without a doubt feasible for any police department and crime analysis unit regardless of financial resources. There is no doubt that specialized commercial products have and will continue to play an important role in police departments that can afford them, but sophisticated, high end and free software products are available that can produce forecasts rivaling or at least coming close to many expensive commercial products. So for many agencies it makes sense to invest in training and, since data is such an important factor in producing good crime forecasts, to ensure that their crime analysis units have the most complete, accurate and wide ranging data sets available.

Dr. Andreas (Olli) M. Olligschlaeger specializes in law enforcement information systems consulting and software development for federal, state and local agencies. Formerly a systems scientist at Carnegie Mellon University, with appointments at the H. John Heinz III School of Public Policy, the Robotics Institute and the School of Computer Science, Olli also has practical experience working with law enforcement agencies in narcotics enforcement, crime analysis and criminal intelligence. The primary focus of his work is on artificial intelligence methods for crime forecasting, advanced analytical tools for the automated mining of very large data sets for both crime analysis and criminal intelligence, advanced spatial statistical methods for geographic information systems and crime mapping, and the development of law enforcement related systems that integrate many different analytical techniques into a single interface. His work in correctional intelligence has included the automated mining of and topic detection within speech recognized text derived from inmate telephone calls. Olli is a member of the International Association of Crime Analysts, the International Association of Law Enforcement Intelligence Analysts, the Society of Police Futurists International, where he is the immediate past president, the PFI/FBI Futures Working Group and serves on the advisory board of the High Tech Crime Consortium.
References


Data and Research for Increased Safety and Fairness

By Anne Milgram

Over the past two decades, crime rates in the United States—and, in particular, violent crime rates—have declined sharply, and leading experts suggest that innovative policing strategies deserve a large part of the credit for this decrease. From CompStat to hot-spotting, law enforcement has led the way in leveraging data and technology to determine how best to deploy its resources to enhance public safety.

And yet significant challenges remain. Violent crime may be down overall, but it continues to plague many urban neighborhoods. Recidivism is disturbingly high, with nearly two-thirds of people who leave jail or prison winding up back behind bars within three years. And this is despite a dramatic increase in spending: The states and the federal government now spend $70 billion a year on corrections, an increase of 660 percent since 1982.

So how can we change this? While serving as Attorney General of New Jersey from 2007 to 2010, I spent a great deal of time considering ways we could make our criminal justice system more effective at reducing crime and producing just outcomes. For the past three years, I have continued to tackle those same issues from a different angle—as the Vice President of Criminal Justice at the Laura and John Arnold Foundation (LJAF).

What I have come to believe based on my experience in both posts is that the key to improving public safety, fairness, and cost-effectiveness is the use of data and research to guide criminal justice policies and decisions. In everything from education to health care, this isthe approach that has been successful in addressing many societal challenges. Clearly, we can—and must—do the same in the field of criminal justice.

My colleagues and I at LJAF have spent much of the past three years gathering data, conducting research, and developing tools to help police, prosecutors, judges, and others address some of the key challenges they face every day.

One of the first questions we focused on is how to determine—at the earliest stages of the criminal justice process—what risks a given defendant poses to public safety. This, of course, is a question police must grapple with as they make decisions about who to arrest, who to cite, and who simply to give a warning. Likewise, prosecutors face this issue when making decisions about charge, diversion, deferred prosecution, and recommendations to the court about how to deal with a defendant while his case

Anne Milgram, JD, is the Vice President of Criminal Justice at the Laura and John Arnold Foundation.

The Laura and John Arnold Foundation (LJAF) is a private foundation committed to producing substantial, widespread, and lasting reforms that will maximize opportunities and minimize injustice in our society. Its strategic investments are currently focused in criminal justice, education, public accountability, and research integrity. LJAF has offices in Houston and New York City.

www.arnoldfoundation.org
is ongoing. And judges, too, must assess risks to public safety when they decide whether—and under what conditions—to release a defendant before trial.

Yet, despite the importance of these decisions, very few jurisdictions use research-based instruments to assist those making these critical determinations about risk. Unfortunately, this means that dangerous defendants are often back out on the street quickly, while significant numbers of low-risk defendants remain incarcerated for long periods.

To address this problem, we are building data-driven tools that provide police, prosecutors, and judges with objective information about the risk that a defendant will commit a new crime, commit a new violent crime, or fail to appear for court.

The first of these instruments, the Public Safety Assessment -Court (PSA-Court), is already in use in pilot sites across the country. The instrument was developed using the largest-ever dataset of 750,000 cases from across the country, which were rigorously analyzed to determine what information about a defendant is most predictive of risks to public safety. Based on this research, we created the PSA-Court, which relies on just nine objective risk factors—all of which can be gathered from the administrative record without interviewing the defendant—to determine the likelihood that a defendant will be rearrested, engage in violence, or skip court.

Kentucky, whose judges have been using the PSA-Court since July 2013, has seen striking results: In the first six months of use, judges have been able to reduce pretrial crime by 15 percent, while at the same time increasing the proportion of defendants released before trial. And the PSA-Court is providing them with critical information about violence. In Kentucky, defendants flagged by the assessment as posing an increased risk of violence are, in fact, rearrested for violent acts at a rate 17 times that of defendants who are not flagged.

Importantly, because it does not rely on factors like neighborhood or income, the PSA-Court is helping deliver these results without discriminating on the basis of race or gender.

Soon, we will begin piloting a similar tool for prosecutors, the PSA-Prosecution, which will help inform their decisions regarding charge, diversion, and deferred prosecution.

And on the policing front, we are partnered with the International Association of Chiefs of Police to conduct research on how police departments currently use citations in lieu of arrest. This work will provide the foundation for developing instruments to help law enforcement make data-driven decisions about which individuals pose a risk of committing a new crime or failing to come back to court, and therefore should be arrested and booked rather than cited and released.

But risk assessment is not the only area in which data and technology can help improve public safety. LJAF has also developed an app for tablets that will allow police to conduct eyewitness identifications quickly and effectively in the field. It will also provide law enforcement agencies with a simple way to adopt research-based best practices that improve the accuracy of identifications. Given the important role that eyewitness IDs play in investigations and trials, enhancing their accuracy can help get dangerous people off the streets.

From policing to prosecution to courts, data and research can generate powerful insights and help develop tools that make our neighborhoods safer, promote fair outcomes, and to help use our resources as effectively and efficiently as possible.
Foreclosures, Domestic Disturbances, and Policy Implications

By Kim Lersch, Paul Cromwell, and Christine Sellers

Over the past decade, the United States has experienced a foreclosure crisis that reached epic proportions, rivaling the level seen during the Great Depression of the 1930s. While nationally the foreclosure activity was at a 5 year low in the third quarter of 2012, some states have continued to see their rates climb year after year. In May 2014 the national foreclosure rate for the entire U.S. was 1 in every 1,199 housing units. The State of Florida reported the highest rates in the country with 1 in every 436 homes in some stage of foreclosure. Conversely, the rate in North Dakota was 1 in every 160,000 homes (RealtyTrac, 2014).

Foreclosures and Domestic Issues

While the popular media and trade publications have focused attention on a foreclosure – crime link, the scholarly literature on the impact of foreclosure on crime and disorder is just beginning to emerge. One area that has not been addressed is the association between foreclosures and crimes of a distinctly domestic nature. The extant research suggests that the foreclosure rate does have an impact on the community, but it may also negatively impact families because of the financial stress that foreclosure often reflects (Ross & Squires, 2011). A small body of self-report survey research has linked economic hardship to marital dissatisfaction and instability (Conger, Elder, Lorenz, Conger, Simons, Whitbeck, Huck, & Melby, 1990), family conflict (Fox & Chancey, 1998), and intimate partner violence (Fox, Benson, DeMaris, & Van Wyk, 2004; Weissman 2007). The purpose of the present study is to explore the relationship between calls for service involving violent and non-violent domestic disturbances and foreclosure filings in the City of Tampa, FL.

The Tampa Experience

In 2008, the Tampa Police Department responded to a total of 16,773 calls for service classified as “domestic disputes” by the 911 call center. There were 3,206 calls that specifically included allegations of violence, use of a weapon, or aggravated assault (see Figure 1).

Foreclosure data was obtained through ForeclosuresDaily.com for fiscal year 2008. There were 5,322 foreclosure filings on residential properties within the jurisdictional bounds of the Tampa Police Department (TPD). To determine whether or not there was a relationship between the foreclosure locations and the calls for service we first examined the density of the locations of residential foreclosures and the concentration of calls for service for domestic disturbances (See Figure 2). Kernel density map layers were created and classified according to standard deviation breaks. Visual inspection of these two density maps indicates similar patterns of concentration of points.

Kim Lersch is currently the Director of the School of Public Affairs at the University of South Florida. She received her PhD from the University of Florida in Sociology in 1995 and a graduate certificate in GIS from University of South Florida in 2010. Her research interests include the spatial distribution of crime, and planning for crime prevention. Currently Dr. Lersch and co-author Tim Hart are working on the 4th edition of *Space, Time, and Crime*. 

www.crimemapping.info  www.policefoundation.org
is ongoing. And judges, too, must assess risks to public safety when they decide whether—and under what conditions—to release a defendant before trial.

Yet, despite the importance of these decisions, very few jurisdictions use research-based instruments to assist those making these critical determinations about risk. Unfortunately, this means that dangerous defendants are often back out on the street quickly, while significant numbers of low-risk defendants remain incarcerated for long periods.

To address this problem, we are building data-driven tools that provide police, prosecutors, and judges with objective information about the risk that a defendant will commit a new crime, commit a new violent crime, or fail to appear for court.

The first of these instruments, the Public Safety Assessment -Court (PSA-Court), is already in use in pilot sites across the country. The instrument was developed using the largest-ever dataset of 750,000 cases from across the country, which were rigorously analyzed to determine what information about a defendant is most predictive of risks to public safety. Based on this research, we created the PSA-Court, which relies on just nine objective risk factors—all of which can be gathered from the administrative record without interviewing the defendant—to determine the likelihood that a defendant will be rearrested, engage in violence, or skip court.

Kentucky, whose judges have been using the PSA-Court since July 2013, has seen striking results: In the first six months of use, judges have been able to reduce pretrial crime by 15 percent, while at the same time increasing the proportion of defendants released before trial. And the PSA-Court is providing them with critical information about violence. In Kentucky, defendants flagged by the assessment as posing an increased risk of violence are, in fact, rearrested for violent acts at a rate 17 times that of defendants who are not flagged. Importantly, because it does not rely on factors like neighborhood or income, the PSA-Court is helping deliver these results without discriminating on the basis of race or gender.

Soon, we will begin piloting a similar tool for prosecutors, the PSA-Prosecution, which will help inform their decisions regarding charge, diversion, and deferred prosecution.

And on the policing front, we are partnered with the International Association of Chiefs of Police to conduct research on how police departments currently use citations in lieu of arrest. This work will provide the foundation for developing instruments to help law enforcement make data-driven decisions about which individuals pose a risk of committing a new crime or failing to come back to court, and therefore should be arrested and booked rather than cited and released.

But risk assessment is not the only area in which data and technology can help improve public safety. LJAF has also developed an app for tablets that will allow police to conduct eyewitness identifications quickly and effectively in the field. It will also provide law enforcement agencies with a simple way to adopt research-based best practices that improve the accuracy of identifications. Given the important role that eyewitness IDs play in investigations and trials, enhancing their accuracy can help get dangerous people off the streets.

From policing to prosecution to courts, data and research can generate powerful insights and help develop tools that make our neighborhoods safer, promote fair outcomes, and help use our resources as effectively and efficiently as possible.
However, if one examines the northernmost area of the TPD jurisdiction, there are areas with high concentrations of residential foreclosures with no corresponding clustering in calls for domestic disturbances. This area of the City, known as New Tampa, is somewhat unique. Over the past decade this area experienced a great deal of growth as developers constructed a number of affluent, gated communities.

We then wished to determine whether or not the rate of foreclosure filings occurring within a census tract was a significant predictor of the rate of domestic disturbance calls when controlling for socioeconomic variables. GeoDa was used to construct a spatial lag model. The foreclosure rate continued to be a significant predictor of the rate of domestic disputes, even when controlling for concentrated disadvantage, household size, and other demographics. The model was able to explain nearly 75% of the variation in the rate of domestic disputes (for a more detailed discussion of the methodology and results, please see Lersch, Sellers, & Cromwell, in press).
Policy Implications

The results of this exploratory study hold important policy implications for law enforcement agencies and social service providers. Our findings suggest that high levels of foreclosure filings may be related to heightened stress, which ultimately could manifest itself in escalated levels of violence and other domestic disputes and increasing calls for police intervention.

Perhaps the most important policy issue suggested by this research is how best to respond to domestic disturbance calls for police service and whether alternative criminal justice processes should be considered by policy makers?

Historically, domestic disturbance calls for service were treated as a family matter not appropriate for police intervention. The criminal justice system preferred to let the parties “work out” their problems privately. Studies of police encounters in domestic violence situations have found that arrest was not the most common response. The range of responses by police included ordering participants to cease their illegal behavior, separating victim from abuser, counseling the participants, referring participants to social service agencies, and arresting one or both participants. Arrest was not the most common outcome. Studies found that arrest rates ranged from 12-40% (Walker and Katz, 2013).

By the 1970s the women’s movement resulted in political pressure for greater intervention in abuse cases and lobbied for mandatory arrest policies (Flannigan, 2013; Weissman, 2007). In the 1980s, prosecutors in some jurisdictions began to develop special programs for domestic violence cases. These policies tended to focus on police arrest practices; in many jurisdictions, requiring mandatory arrest for abusers. Prosecutors were encouraged to institute “no drop” practices, which required prosecution even without the abuse victim’s cooperation, and courts were also urged to impose more severe sanctions (Weissman, 2012).

But, how did these criminal justice focused policies affect the incidence of domestic violence? Recent findings by the U.S. Department of Justice found that the rate of intimate partner violence fell 42-49% from 1993-2005 (Bureau of Justice Statistics, 2013); however, unofficial data from social services agencies and organizations such as the National Domestic Violence Hotline (NDVH) have reported increases in incidents of domestic violence. The NDVH reported that calls were up 21% in September 2008, compared to September 2007. Fifty-four percent (54%) of 7,865 of NDVH callers who agreed to participate in a survey reported that there was a change in their household’s financial situation in the past year, and they believed that abusive behavior had also increased during that time period. Notably, this data corresponds to the beginning of the economic downturn and housing crisis.

The disparity between “official” statistics and calls for social services assistance may also be partially explained by the level of police response. Although most research has shown that arrests reduce the incidence of repeat violations, others suggest that mandatory arrest discourages reporting (Weissman, 2012; Walker and Katz, 2013).

Domestic violence victims may instead seek help from non-law enforcement sources, and their victimization may not be reflected in police calls for service statistics. We suggest that vigorous law enforcement tactics may reduce reporting in some cases. A woman who is financially dependent on her abuser and obliged to live with him may not report her abuse. Victims who are in this country illegally may fear deportation if they...
report their abuser. Lesbians, gay men, and transgendered victims may fear discrimination, and African-American, Hispanic, and persons in communities with a history of distrust of police may also be averse to involving authorities. Women may also fear further violent victimization when their abuser is released from custody (Weissman, 2007; Maxwell and Stone, 2010). Furthermore, Robinson and Welchans (2003) reported that victims in middle and upper-middle class neighborhoods may fail to report through official channels due to embarrassment or concerns over loss of income if the spouse is arrested. While this is purely speculative, it may be that in the more affluent New Tampa area there were incidents of domestic disturbances that simply were not reported to the police.

This data suggests that mandatory arrest, “no drop” prosecution, and other vigorous law enforcement and prosecutorial practices, while focused on concern and safety for the victim, may be too narrowly focused. We suggest that response to domestic violence should include a range of strategies, including police and criminal justice interventions. We agree with Weissman (2007) that a return to the “hands-off” approach is inconceivable.

The present study found that the rate of foreclosure filings was a significant predictor of the rate of calls for service for domestic disturbances. Law enforcement agencies and social service providers may wish to target their scarce resources to enhance domestic violence prevention efforts in areas with high levels of actual or potential foreclosure filings. Revised police response policies providing specialized training for patrol officers in the handling of participants in domestic disputes and violence should be implemented in each department. Special domestic violence units could employ a range of strategies which include arrest, but also look toward prevention, referral to appropriate social service agencies, emergency shelters, employment, and financial counseling when the circumstances are appropriate.
References


**Using Risk Terrain Modeling Technique to Identify Places With the Greatest Risk for Violent Crimes in New Haven**

By Charles Anyinam

**Introduction**

The importance of place in crime causation and crime prevention has been emphasized over the years (Weisburd, 2008). Current environmental criminological theories of crimes such as rational choice (Clarke & Felson, 1993); routine activity theory (Cohen & Felson, 1979); and crime pattern theory (Brantingham & Brantingham, 1993) give weight to the importance of place for understanding crime. As McCord and Ratcliffe have summed up, “together these three theories state that specific types of land uses and facilities generate crime due to the daily activities associated with them and the number and types of people they attract” (McCord & Ratcliffe, 2009, p. 18). With increasing effectiveness, GIS and spatial analysis are being used to more rigorously examine the effect of “place” on crime. In the last few years, one approach that has been advanced to understand the environmental context of crime and the role that sites, locations or places in urban areas play in attracting criminal activities is Risk Terrain Modeling (RTM) (Caplan & Kennedy, 2010a; Caplan, Kennedy, & Miller, 2011).

Risk Terrain Modeling is an analytic technique based on the idea that crime offenders, crime victims, and crime targets operate in space and time and that the risk of a crime event occurring at a specific location is determined by a combination of social, cultural, economic, and physical environmental risk factors (Caplan & Kennedy 2010; Caplan & Kennedy, 2011; Caplan, Kennedy & Piza 2012). When these risk factors, which may be “generators” and/or “attractors” of crime, intersect in space and time, they have the greatest potential to yield a particular crime outcome. When such factors are identified, they can be used to create separate map layers to represent their presence, absence, or intensity (Caplan & Kennedy, 2010). When these risk map layers are further combined, a composite “risk terrain” map can be produced. Risk terrain modeling of crimes, thus, results in maps that show places with the greatest risk or likelihood of becoming places for crimes to occur in the future. The result is the production of a geospatial hot spot map of areas with the highest probability of crimes occurring within a certain time period. This analytic technique provides crime analysts the opportunity to describe, demonstrate, analyze and explain the socio-cultural, economic, and physical/geographical environmental context within which certain types of crime are concentrated in urban areas. In a nutshell, the RTM technique examines the features of places that contribute to crime concentration.

Charles Anyinam holds a PhD (Geography) from Queen’s University, Kingston, Ontario, Canada and a Graduate Diploma (GIS) from York College of Information Technologies, Toronto. He taught at a number of universities in Canada including University of Toronto and York University, North York, Ontario, Canada. He is currently the Supervisor of the Crime Analysis Unit, New Haven Police Department, New Haven, Connecticut. His research interests include spatial-temporal analysis, predictive analytics, and use of a variety of techniques in crime mapping and analysis.
This report briefly illustrates how the risk terrain modeling (RTM) technique was used to identify the environmental context within which violent crimes are most likely to occur in the city of New Haven, Connecticut. The report also tests the predictive value of the composite “risk terrain” map to determine the empirical credibility and degree of confidence in using the model for future forecasts of violent crimes in New Haven.

Identifying Risk Factors Using RTMDx Utility Software

New Haven, Connecticut is the second-largest city in Connecticut with a population of 129,779 people in 2010. The city has a total area of 20.1 square miles (52.1 km²), of which 18.7 square miles (48.4 km²) is land and 1.4 square miles (3.7 km²), or 6.67%, is water. Several risk factors that have been found to relate to violent crimes by existing criminological theories and empirical studies are also found to be relevant in the city of New Haven. Based on personal knowledge of criminal activities in New Haven as well as past literature (for example, Anderson, 1999; Bernasco & Block, 2011; Brantingham & Brantingham, 1982, 1995; Caplan et al., 2011, Caplan, Kennedy, & Baughman, 2012; Fass & Francis, 2004; McCord, Ratcliffe, Garcia, & Taylor, 2007), 27 potential risk factors were selected as “attractors” and/or “generators” of violent criminal activities in New Haven. For convenience, these are grouped into three main categories:

A. Commercial Infrastructural Services

Banks, bars/clubs/restaurants, beauty salons/barber shops, check-cashing stores, convenience stores, entertainment facilities, fast foods, gas stations, hotels/motels/inn, package stores, pawn shops, recreational facilities, strip malls

B. Municipal Infrastructural Services

Apartment complexes, bus stops, cemeteries, elderly housing, halfway houses, parks, public housing, schools, Section 8 housing

C. Potential Offenders/Perpetrators of Crime

Drug arrestees’ home addresses, public drug complaint calls for service, parolees, released prisoners, probationers

The Risk Terrain Modeling process tests a variety of factors that are thought to be geographically related to crime incidents. Valid factors are selected and then weighted to produce a final model that basically paints a picture of places where crime is statistically most likely to occur (Caplan, et al. 2011). In 2013, Rutgers Center for Public Security released the RTMDx Utility software, a tool that facilitates the empirical method of risk terrain map production. Prior to the development of this tool, the operationalization of risk factors was undertaken manually using the 10 steps described by Caplan and Kennedy (2010) with the help of an ArcGIS toolbox designed by Caplan and colleagues for use in ArcMap.

The 27 factors identified above were operationalized using the Risk Terrain Modeling Diagnostics (RTMDx) software that automates various steps required to determine the spatial influence and significance testing of risk factors (Caplan et al., 2013). The RTMDx tool accepts up to 30 risk factors as inputs for analysis. The 27 identified risk factors were operationalized by inputting a number of parameters: shape of study area, block length, raster cell size (i.e. the dimensions on the ground of a single cell in a raster, measured in map units), type of model, outcome event, risk factors and operationalization of the spatial influence (Caplan et al., 2013). The following are inputs set up before running the
RTMDx software:

1. **Study area:** New Haven city boundary.

2. **Cell Size:** the average block length in New Haven is 446 feet and the raster cell size was set to half-block (223 ft). There were 11,335 raster cells used in the analysis of which 1984 cells contained violent crime events.

3. **Model Type menu allows one to select either aggravating or protective model type.** An aggravating model assumes that the risk factors input into the RTMDx utility correlate with the locations of outcome events and it tests for positive spatial relationships. A protective model type assumes that the risk factors correlate with the absence of outcome events and it tests for negative spatial relationships. One example of an aggravating factor might be ATMs for the crime of robbery. An example of a protective factor could be a police sub-station or community garden (Caplan et al. 2013, p.18). Aggravating model was input in as the model type for this study.

4. **Outcome Event Data:** The outcome events for the study are violent crimes (principally murders, aggravated assaults and robberies) that occurred in New Haven between 2009 and 2013.

5. **Risk Factors and Operationalization of Spatial Influence:** 27 risk factors were individually added, each of which was spatial influence-based and tailored to the study area: density, proximity, and both proximity and density. In this study, the maximum spatial influence was set to three blocks for each risk factor with spatial increments set to half-block. Caplan, et al. (2011) have demonstrated that some risk factors are more a function of distance from the closest feature while others are a function of the presence or absence of the feature (i.e. density of the feature). In this study, drug complaints data for 2009-2013, public housing data, drug arrestees’ home addresses, prisoners released to the city of New Haven in 2013, Section 8 houses, apartment complexes, and all 2013 parolees and probationers were operationalized as a function of density. The rest were operationalized as a function of proximity.

The RTMDx tool was run to generate a model that represented the risk factors for 6,671 violent crime events. Based on the operationalization of the 27 risk factors, 162 variables were created and tested for significance. The software builds from a null model and uses bidirectional stepwise regression process to build the “best model”. RTMDx operates based on building the optimal model, which is reflected in the Bayesian Information Criteria (BIC) value. The “Best Model Specification” section of the report generated after running the RTMDx tool provides details about the risk factors included in the risk terrain model, their optimal spatial influences and operationalization methods and their relative weights (Caplan et al., 2013).

The RTMDx tool determined that the best risk terrain model was a Negative Binomial type II model with 16 of the 27 risk factors and a BIC score of 14,067. The model also included an intercept term that represents the background rate of events and an intercept term that represents over-dispersion of the event counts. The 16 risk factors selected after running the software make statistically significant contributions to the outcome variable. Thus, all areas in New Haven with a high concentration of drug complaints, public housing, convenience stores, gas stations, bus stops, drug arrestees’ home addresses, residences of released prisoners, banks, bars, restaurants and cafes, schools, section 8 houses, apartment complexes, probationers’ residence, beauty salons and barber shops, package stores, and parks are associated with a higher concentration of violent crimes in the city (Fig. 1).
New Haven, Connecticut is the second-largest city in Connecticut with a population of 129,779 people in 2010. The city has a total area of 20.1 square miles (52.1 km²), of which 18.7 square miles (48.4 km²) is land and 1.4 square miles (3.7 km²), or 6.67%, is water. Several risk factors that have been found to relate to violent crimes by existing criminological theories and empirical studies are also found to be relevant in the city of New Haven. Based on personal knowledge of criminal activities in New Haven as well as past literature (for example, Anderson, 1999; Bernasco & Block, 2011; Brantingham & Brantingham, 1982, 1995; Caplan et al., 2011, Caplan, Kennedy, & Baughman, 2012; Fass & Francis, 2004; McCord, Ratcliffe, Garcia, & Taylor, 2007), 27 potential risk factors were selected as “attractors” and/or “generators” of violent criminal activities in New Haven. For convenience, these are grouped into three main categories:

A. Commercial Infrastructural Services
   - Banks, bars/clubs/restaurants, beauty salons/barber shops, check-cashing stores, convenience stores, entertainment facilities, fast foods, gas stations, hotels/motels/inns, package stores, pawn shops, recreational facilities, strip malls

B. Municipal Infrastructural Services
   - Apartment complexes, bus stops, cemeteries, elderly housing, halfway houses, parks, public housing, schools, Section 8 housing

C. Potential Offenders/Perpetrators of Crime
   - Drug arrestees’ home addresses, public drug complaint calls for service, parolees, released prisoners, probationers

The Risk Terrain Modeling process tests a variety of factors that are thought to be geographically related to crime incidents. Valid factors are selected and then weighted to produce a final model that basically paints a picture of places where crime is statistically most likely to occur (Caplan, et. al 2011). In 2013, Rutgers Center for Public Security released the RTMDx Utility software, a tool that facilitates the empirical method of risk terrain map production. Prior to the development of this tool, the operationalization of risk factors was undertaken manually using the 10 steps described by Caplan and Kennedy (2010) with the help of an ArcGIS toolbox designed by Caplan and colleagues for use in ArcMap.

The 27 factors identified above were operationalized using the Risk Terrain Modeling Diagnostics (RTMDx) software that automates various steps required to determine the spatial influence and significance testing of risk factors (Caplan et al., 2013). The RTMDx tool accepts up to 30 risk factors as inputs for analysis. The 27 identified risk factors were operationalized by inputting a number of parameters: shape of study area, block length, raster cell size (i.e. the dimensions on the ground of a single cell in a raster, measured in map units), type of model, outcome event, risk factors and operationalization of the spatial influence (Caplan et al., 2013). The following are inputs set up before running the RTMDx.

FIG. 1 shows results of the “best model” specified after running RTMDx. The relative risk value column shows the weighted values of the selected risk factors. Ten of the risk factors were operationalized based on proximity function of half-block (223 feet). The top five relative risk values are associated with drug complaints, public housing, and location of convenience stores, gas stations and bus stops. Using the “best model” for creating high-risk areas, the RTMDx creates a composite risk terrain map.

### Table

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Operationalization</th>
<th>Spatial Influence</th>
<th>Coefficient</th>
<th>Relative Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>DrugComplaints2009 2013</td>
<td>Density</td>
<td>223</td>
<td>1.3281</td>
<td>3.7739</td>
</tr>
<tr>
<td>Rate</td>
<td>PublicHousing</td>
<td>Density</td>
<td>223</td>
<td>0.9946</td>
<td>2.7037</td>
</tr>
<tr>
<td>Rate</td>
<td>ConvenienceStores</td>
<td>Proximity</td>
<td>1338</td>
<td>0.9262</td>
<td>2.5249</td>
</tr>
<tr>
<td>Rate</td>
<td>GasStations</td>
<td>Proximity</td>
<td>223</td>
<td>0.7288</td>
<td>2.0726</td>
</tr>
<tr>
<td>Rate</td>
<td>BusStops</td>
<td>Proximity</td>
<td>223</td>
<td>0.6148</td>
<td>1.8492</td>
</tr>
<tr>
<td>Rate</td>
<td>DrugArrestesHomeLocation</td>
<td>Density</td>
<td>1115</td>
<td>0.5965</td>
<td>1.8157</td>
</tr>
<tr>
<td>Rate</td>
<td>PrisonersReleased2013</td>
<td>Density</td>
<td>223</td>
<td>0.5274</td>
<td>1.6945</td>
</tr>
<tr>
<td>Rate</td>
<td>Banks</td>
<td>Proximity</td>
<td>223</td>
<td>0.5207</td>
<td>1.6831</td>
</tr>
<tr>
<td>Rate</td>
<td>BarsClubsRestaurantsCafes</td>
<td>Proximity</td>
<td>223</td>
<td>0.4914</td>
<td>1.6347</td>
</tr>
<tr>
<td>Rate</td>
<td>Schools</td>
<td>Proximity</td>
<td>223</td>
<td>0.4201</td>
<td>1.5221</td>
</tr>
<tr>
<td>Rate</td>
<td>Section8Housing</td>
<td>Density</td>
<td>223</td>
<td>0.4180</td>
<td>1.5189</td>
</tr>
<tr>
<td>Rate</td>
<td>ApartmentComplex</td>
<td>Density</td>
<td>223</td>
<td>0.4143</td>
<td>1.5133</td>
</tr>
<tr>
<td>Rate</td>
<td>Probationers2013</td>
<td>Density</td>
<td>1338</td>
<td>0.4072</td>
<td>1.5026</td>
</tr>
<tr>
<td>Rate</td>
<td>BeautySalonBarberShop</td>
<td>Proximity</td>
<td>892</td>
<td>0.2726</td>
<td>1.3134</td>
</tr>
<tr>
<td>Rate</td>
<td>Package Stores</td>
<td>Proximity</td>
<td>1115</td>
<td>0.2422</td>
<td>1.2740</td>
</tr>
<tr>
<td>Rate</td>
<td>Parks</td>
<td>Proximity</td>
<td>446</td>
<td>0.2274</td>
<td>1.2554</td>
</tr>
<tr>
<td>Rate</td>
<td>Intercept</td>
<td>-</td>
<td>-</td>
<td>-2.6195</td>
<td>-</td>
</tr>
<tr>
<td>Overdispersion Intercept</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.7132</td>
<td>-</td>
</tr>
</tbody>
</table>

**Producing A Composite Risk Terrain Map And Assessing Its Predictive Value**

The selected risk terrain model assigned relative risk scores to cells ranging from 1 for the lowest risk cell to 575.9 for the highest risk cell. These scores allow cells to be easily compared. For instance, a cell with a score of 575.9 has an expected rate of crime that is 575.9 times higher than a cell with a score of 1. The RTMDx utility
software (professional version) generates a GeoTiff (i.e. raster image) that contains only values for the cells that were included in the “best model” specifications. The resulting composite map with GeoTiff format was further converted to create a new grid raster layer in order to undertake further analysis of the results.

To determine the effectiveness of the model for forecasting future crime events, the relative risk scores generated by the RTMDx utility tool was used to create top risky places with high probability for violent crimes to occur in the future. The raster grid layer was converted to vector shapefile (Map 1), which was further symbolized using another spatial and visualization technique. For the purposes of operational policing, the Local Indicators of Spatial Auto correlation (LISA) method was used to identify local spatial clusters of similar values (whether high or low values). LISA was applied to the relative risk values derived from the vector shapefile. This process was undertaken with the construction of a spatial weights matrix, which imposes a neighborhood structure on the data to access the extent of similarity between locations and values. Two basic categories of defining neighborhood spatial relation are contiguity and distance. Contiguity-based weights matrices include rook and queen. Areas are neighbors under the rook criterion if they share borders (e.g., on a grid, only the cells to the North-South and East-West are neighbors). Under the queen criterion, areas are neighbors if they share either a border or point (e.g., on a grid, in addition to the four cells included under rook, the four cells sharing a corner with the central location are also counted as neighbors) (Anselin et al. 2008). The queen contiguity was used to identify spatial relationship. The local Moran’s I LISA approach indicated areas of statistically significant clusters where violent crimes are higher or lower than would be expected if the incidents were randomly distributed in New Haven. The advantage of using LISA is that it can distinguish between statistically significant clusters of high values surrounded by high values (HH), low values surrounded by low values (LL), high values surrounded by low values (HL), and low values surrounded by high values (LH) (Kennedy et al. 2011). Map 2 shows results of the LISA cluster analysis that identified only statistically significant clusters of High – High places. These top high-high risk areas make up only 6.09% of the entire area of New Haven.

To what extent does the derived composite risk terrain map in the form of LISA cluster map accurately forecast violent crime locations in New Haven? To find out, the high-high risk areas were tested with new crime data. During the first seven months of 2014, the New Haven police recorded 396 incidents of murder, robbery and aggravated assault (including non-fatal shootings). Specifically, 41% of non-fatal shootings, 57% of murders and 39% of robberies occurred in these high-risk areas, which make up only 6.09% of the city of New Haven.

These findings support the validity of RTM methodology for forecasting. The results of the study demonstrated that the co-existence of the 16 risk factors does have strong effects on the locations of violent crimes in New Haven. The results of the study were discussed with supervisors in charge of the ten policing districts. They were made aware of the high-risk areas and the different risk factors that form the “backdrop” of these areas. The need to formulate appropriate proactive and preventive strategies to address violent crimes in these areas was also discussed. Identifying these high risk areas for violent crimes adds considerable value to not only understanding the nature of violent crime problems in the city as a whole but also in
prioritizing and designing appropriate police and community interventions to address the underlying risks associated with them. The risk terrain model highlights the need to devise, implement and take actions that will deter potential offenders, harden crime-prone targets, and reduce violent crime incidents in the city.

**Conclusion**

In this study, we identified 16 statistically significant risk factors that underlie violent crime occurrences in the city of New Haven, using the RTMDx software recently released by Rutgers Center for Public Security. The results point to the fact that the Risk Terrain Modeling technique has significant operational value. The fact that a high percentage of violent crimes that occurred during the first seven months of 2014 happened in areas identified by the composite RTM map lends support to the need for “place-based” strategies in dealing with violent crimes in New Haven. The results reinforce the relevance and validity of the risk terrain model in identifying geospatial hotspots or clusters of high-risk areas for violent crimes. The high-risk areas identified have the potential to make crime prevention interventions more efficient and successful in terms of resource allocations and short- and long-term planning for crime control and reduction. Risk terrain modeling is a “placed-based” forecasting technique that has the potential to assist police in not only prioritizing areas of focus for action, but also to help patrol officers find more meaningful ways to modify and/or change the geographic characteristics and dynamics that promote, encourage, and cause violence in these risk areas. The RTM approach and tool provides for a better understanding of the need and urgency to address the relatively few high-risk places and to prevent violent crimes in the city of New Haven and elsewhere.
REFERENCES


In this study, we identified 16 statistically significant risk factors that underlie violent crime occurrences in the city of New Haven, using the RTMDx software recently released by Rutgers Center for Public Security. The results point to the fact that the Risk Terrain Modeling technique has significant operational value. The fact that a high percentage of violent crimes that occurred during the first seven months of 2014 happened in areas identified by the composite RTM map lends support to the need for “place-based” strategies in dealing with violent crimes in New Haven. The results reinforce the relevance and validity of the risk terrain model in identifying geospatial hotspots or clusters of high-risk areas for violent crimes. The high-risk areas identified have the potential to make crime prevention interventions more efficient and successful in terms of resource allocations and short- and long-term planning for crime control and reduction. Risk terrain modeling is a “place-based” forecasting technique that has the potential to assist police in not only prioritizing areas of focus for action, but also to help patrol officers find more meaningful ways to modify and/or change the geographic characteristics and dynamics that promote, encourage, and cause violence in these risk areas. The RTM approach and tool provides for a better understanding of the need and urgency to address the relatively few high-risk places and to prevent violent crimes in the city of New Haven and elsewhere.