Analyzing Data for Relationships between Traffic Violations, Automobile Crashes and Crime; A Foundation for the Data Driven Approaches to Crime and Traffic Safety (DDACTS) Model.

Place As a Nexus of Traffic Violations, Automobile Crashes and Crime

There has been some research on the linkages between crashes and crimes[1]. That research has focused only on the individual as the nexus between the two. Using theories from psychology and criminology that research has demonstrated that individuals who commit crimes are also likely to commit traffic violations and ultimately get into collisions. Those individuals are likely to engage in risky behavior whether committing crime or driving a vehicle.

These theories, though, do not adequately explain why they occur clustered in geographical space. They only make the psychological link of behavior to the individual but not the link between the individual and place, especially the frequency at which incidents would have to occur to cluster in time in space to constitute a problem. The number of incidents to be clustered in space and time are not likely to be committed by individuals engaging in all three of these aspects at one time. Instead, it requires an examination of the aspects of place, the context, that attract the same types of people to the same places to allow for the convergence on a regular basis. The examination of place will not only help law enforcement executives determine which areas should receive attention but also what type of countermeasures to enact given the context.

To make the place connection between crashes and crimes there is an integral link between them that needs to be incorporated. Given that accidents likely do not come without illegal driving, traffic violation data becomes important to examine the activity in those places where crashes are occurring.

As the connection of violation, crash and crime analyses in place merges it facilitates the efficient deployment of scarce public service resources are becoming the benchmarks of 21st century policing, law enforcement executives should understand the basic principles, processes and nomenclature of these theories. Analysts should evaluate incidences of each type in the context of criminological and geographical theories that seeks causes for their spatial and temporal clustering and use problem-oriented and information-led policing approaches to develop appropriate analytical reports to inform their command staff.

The DDACTS model ultimately uses spatial analysis to identify areas with disproportionately high incidences of violations, crashes and crime but requires several types of analysis to ensure that the most appropriate places for intervention are identified. The proposed analysis method in this document advocates for a four step sequential process of correlation, spatial, temporal, and ecological analysis.

[1] Brace C. et. al. (2009). *The Relationship Between Crime and Road Safety*. Final Report from Monash University Accident Research Center.

Analytical tools have significantly improved the ability of crime and traffic analysts, along with researchers, to understand incident patterns, as well as patterns of victimization. The use of statistical techniques to relationships of incidents of different types provides firm evidence that both are occurring in the same places and at the same times. This identification of Hot Spots allows law enforcement executives to use High Visibility Enforcement countermeasures to more efficiently impact crime and crashes together.

The extent to which law enforcement agencies are using violation, crash and crime mapping varies greatly, as do the analytical techniques used, the staff involved in the process, and the manner in which the results are used for deployment decisions. In some instances many agencies are not engaging in traffic analysis of any kind. In spite of this variation, as more law enforcement agencies adopt DDACTS and other data-driven approaches for community law enforcement and problem solving, the need for trained personnel and the importance of advanced analytics will grow steadily. Ultimately, the usefulness of the results rests with the proficiency of the individuals using it and the quality of the data used.

Analysis of Data

To adequately identify where high visibility enforcement for the DDACTS model, law enforcement executives must be prepared to track traffic violation, crash and crime data from the entire jurisdiction. This allows for the identification of a number of sites that might be the best candidates for an enforcement program. Further, comparisons among areas in which high visibility enforcement is and is not conducted can be made, as well as at different times. The information below addresses some basic considerations for analyzing data to implement the DDACTS model. It includes preliminary details on the use of correlation, spatial clustering, temporal, and contextual analysis techniques for identifying and analyzing clusters of simultaneous occurrence of incidents in particular places.

Base Data

Since crash frequencies are highly variable from year to year (Nicholson, 1985[2]; 1986[3]), police departments should use at least three (3) years of data to establish a baseline for identifying the locations where violations, crashes and crimes are occurring simultaneously. The use of a single year of data for identifying these locations will appropriately identify places where trends are emerging or are currently in existence but may yield misleading results due to not enough time having passed to ensure there are long term trends. Therefore, multiple years is needed to establish long term trends.

[2] Nicholson, A. J. (1985). *The Variability of Accident Counts*. Accident Analysis and Prevention: 17(1), 47-56.
[3] Nicholson, A. J. (1986). *The Randomness of Accident Counts*. Accident Analysis and Prevention: 18(3), 193-198.

These data should be partitioned at multiple units of times to examine for short-term activity and long-term emergent trends. Data should be saved in 1, 3 and 6 month, as well as 1 year time frames for analysis. This allows for analysis to be done on monthly, quarterly, semi-annual and annual time frames. Crash data should be categorized into the following:

- Property vs. Fatal; and
- Intersection vs. Street.

Crash data should have incidents on the highway removed. This is because the nature of the travel involved and the type of roads accidents occur on they will not contribute to the understanding of any relationship that might be present between the two. The incidents to be analyzed should be logically connected through the environment they are in.

Crime data should be categorized into the Uniform Crime Reporting (UCR) format of:

- Homicide;
- Robbery
  - Commercial
  - Residential
- Aggravated Assault
- Burglary
  - Commercial
  - Residential
- Larceny – Theft
- Motor Vehicle Theft

While the main focus of the analysis will be on crashes and crime it is necessary to link them with the violations. Violations, crashes and crimes should be broken down by type in order to analyze the spatial and temporal correlations between them. Violation data should be categorized to generally represent the following:

- Place holder until determined;

Geographical Units of Analysis

Analysts should select small geographical units for analysis; the preferred unit being the block group if an administrative unit be used and a grid if a standardized unit can be used. An administrative unit will allow for some degree of correlation between crashes and crimes, given they don't occur in the exact same space. Additional analysis units to consider include traffic analysis zones, police beats, or other administrative units, as appropriate. A grid will allow for the standardization of units to minimize problems that arise from using administrative units.

There are two reasons for using small geographical units. First, since most crashes occur on roads and most crimes occur either on sidewalks or within a property boundary (parcel), exact locations will rarely coincide. Second, common factors are likely to involve the interaction between the road system and the land uses it traverses.

As a recommendation it is idea that a grid be used to for analysis. This is because the grid contains units (cells) that removes the Modifiable Areal Unit Problem (MAUP) specifically the zonal problem. The zonal problem arises in administrative units of analysis in that they are often created following streets for compactness or efficiency. This leads to units of different shapes and a wide variation in the number of boundaries it shares with neighbors that, if changed, can affect analysis results because of the different areas they represent. Further, these boundaries can be of varying sizes that can capture more or less data simply because of their size. To alleviate this problem it is recommended that the grid cell size be created that are an artifact of some administrative unit of analysis that contains data about the demographics or aspects of what is being studied and the shape and size characteristics of the jurisdiction. This is because the jurisdiction itself can suffer from the MAUP in that the reduction or enlargement of the jurisdiction can produce different analysis results.

A method has been developed that can identify a grid cell that takes into account an administrative unit and the jurisdiction to produce a unit that preserves compactness but scales to the size of jurisdiction allowing. Furthermore a method has been developed that can partition administrative unit data to those grid cells for analysis.

Analysis of Crimes and Crashes

To be effective, several types of analysis must be conducted to systematically identify the places which are target candidates for DDACTS enforcement. Given that the nexus of crimes and crashes are is place, one of the main and account for the type of crime or crash, its location, and the time of day it occurred. Knowing whether a clusters have an abundance of DWI crashes, auto thefts, and robberies that occur mostly in the evening, as opposed to other types of crashes and burglaries that occur mostly in the afternoon, will greatly influence the deployment of High Visibility countermeasures.

While there may be no causal link between the two it is important to understand that the benefits of an enforcement strategy from DDACTS can be realized none-the-less. This is based on the premise that police presence and activity can deter not just traffic problems but crime (cites for the diffusion of benefits).

Correlation Analysis

Crime, Traffic Violation and Crash data should be aggregated to an appropriate, and if possible the smallest unit of analysis that makes sense and analyzed for correlation

between the three broken down by type. This is the first step in not only filtering data for further analysis but also for identifying those crimes, traffic violations and crashes that appear to have some relationship with each other. This does not mean that there is a causal link, but a geographic link that can guide further analysis for more detailed and substantive linkages.

- Given that there is an enormous amount of data and possibilities for analysis it is imperative to filter data out to focus on those relationships that are have significance. Correlation tables should be constructed to examine the strength of the associations between the crime, violation and accident types.

- The correlation combinations that are possible across the three categories of data and their breakdowns into specific types pare down to:

    - Violations to Crashes but not Crime;
    - Violations to Crime but not Crashes;
    - Crashes to Crime but not Violations;
    - Violations to Crashes and to Crime; and
    - No correlation between Crime, Violations nor Crashes.

  Relationships between these categories are indicators of different problems that the jurisdiction faces and need to be explored for the places where they are occurring.

  A matrix should be developed of what the correlations between, or absence of, incidences might indicate to an analyst.

	Violations	Crashes	Crimes
Violations			
Crashes			
Crimes			

Correlation analysis does not yet isolate the places where these correlations exist but establishes the focus of those crime, violation and accident types for more detailed analysis.

Spatial Analysis

As a next step, the analysis of those crime, violation, and crash types that have been identified through correlation need to be identified with the places that they are occurring. Given that correlations are based on aggregation to areal units and calculated together likely indicates that there are some places that have high volumes of those types and others other units did not. That, however, does not mean that they are all causally related. It is now important to identify which places have these correlations.

Spatial clustering techniques should be used to identify clusters of crashes, violations, and crime types that overlap. As a first step, they will need to identify the degree of global spatial autocorrelation necessary to consider an area a Hot Spot. This is necessary because, oftentimes, the areas in which crimes occur that are more concentrated than those areas in which violations occur. This is the same with crashes. It also extends to five of the combinations listed above of the relationships between each category. A category may be highly concentrated relative to the population distribution. Specific Hot Spots should be identified for crashes and crime types independently. There are a number of techniques for performing this analysis.

- Employing a Kernel Density Estimation (KDE) technique, a visualization of the concentration of events can be conducted using a fixed bandwidth (standard search distance) to identify clusters of crimes and crashes. This allows for the scale of identified clusters to be consistent for comparative purposes. Since crashes are confined to a street network, the standard for a fixed distance should be relatively small. A smaller fixed distance will result in the identification of a more precise area in which to implement high visibility countermeasures.

  But to use KDE it requires identifying the appropriate distance at which to find those clusters.

- Ripley's K in *CrimeStat*[4] can be used to identify a fixed distance that represents the strength at which spatial relationships wane. A quartic function is commonly used to identify the fixed distance, because it is more compact and only considers those observations that fall within the specified fixed distance for clustering. When used for all crime types and crashes, the average distance between a crime type and crashes will become the fixed distance to use. Given that the size of the bandwidth will be small, and likely non-normal, the program provides a distance decay weighting that falls off systematically in calculating estimates that are more uniform under the kernel.

- but it is recommended that analysts use the Nearest Neighbor Hierarchical Clustering (NNHC) routine in *CrimeStat.* This is because it identifies clusters of joint relationships between two variables.

  NNHC identifies clusters of incidents that are closer together than random chance. There are two types of geographic outputs from the NNHC technique, standard deviational ellipses (SDE) and convex hulls. The convex hulls should be used for comparison between crime types and crashes, since they are more precise than the SDEs for identifying true geographic distribution. Based on the prevalence of a crime type and specific crash types, comparisons should focus on time of day. Once identified, analysts should overlay each of the results from the crime and crash incidents. These results can then be used to identify priorities for countermeasure deployment.

[4] http://www.icpsr.umich.edu/CRIMESTAT/

- Univatiate and Multivariate Local Indicators of Spatial Autocorrelation (LISA) can be used to identify clusters of areal units that show clustering of any of the three types of incidences. Univariate LISA simply shows clusters of areas that have high and low values of incidences. It can be valuable to know where there are unusually low incidences of one type, or where there might be a high amount for another type. Multivariate can be used to examine the effect of one type of incident on another for clustering

  Another valuable aspect of using this technique is that it identifies anomalies of high values surrounded by low value clusters and vice ~ versa. This technique can identify areas that statistically appear very different from their adjacent neighbors. These can be

- If data are available, risk-adjusted Hot Spots also should be identified. Crime Hot Spots typically occur where the greatest concentration of people occurs, usually in commercial areas. Crash Hot Spots tend to occur where traffic volumes are highest. To control for the underlying number of persons who could be exposed to crime or crashes, it is preferable to analyze the incidents relative to a baseline of exposure. For crashes, the analysis is the number of crashes relative to vehicle miles traveled (usually in terms of 10 million VMT). For crimes, the analysis is the number of crimes relative to employment or population.

  There are several ways to conduct a risk-adjusted clustering. One way is through a dual kernel density interpolation that interpolates crashes or crimes to small grid cells followed by the interpolation of VMT or employment. A second way is to conduct risk analysis through the risk-adjusted nearest neighbor hierarchical clustering (RNNHC) routine in *CrimeStat*. This routine conducts an NNHC relative to the baseline variable, VMT and employment/population, respectively.

  The identification of Hot Spots using this technique will provide stronger evidence for a concentration of crime and crashes; however, this data can be difficult to obtain on a scale that allows it to be used.

Temporal Analysis

Once the clusters have been identified it will be necessary to examine the temporal distributions of the hot spots as a group and individually to see if the are occurring in the same time frame. Analyzing them as a group will provide an indication if the correlations between crimes and crashes are a metropolitan-wide problem for a particular time of the day. If the results indicate as such this allows for a more strategic approach be taken that distributes policing resources more evenly and takes advantage of more regular activity. Analyzing them individually allows for the identification of specific times of day or night that crimes, traffic violations and crashes might be occurring. If they are at different times, then it requires different approaches to solving

the problems of each than simultaneously. If they are at the same time, then this can allow for a simultaneous solution.

Two general time frames that should be examined and compared are:

- Weekday vs. Weekend
- Day vs. Night

Routine activities change dramatically between these two time frames and produce patterns that would likely be masked if not separated out. Of course, there are other options that allow for a more detailed examination, such as daily or patrol shifts.

Once those places have been identified that have spatial and temporal correlations of crimes and crashes attention can now be focused on examining the contextual factors about those areas. From there, targets for enforcement can be established and plans be made.

Ecological Analysis

There is another psychological link that draws individuals of the same personality to place. Given that the linkages between crimes and crashes is place-based it is important to analyze the context for which these incidents are occurring

As such, it is important to conduct the ecological analysis to indentify the potential environmental links between the three problems.

With the possibility that there are different contextual factors that play a role in the places where these simultaneously occur it may be beneficial to tailor countermeasures to best fit the place. Research has shown this type of customization can be effective and there is no reason to believe that it is not the case with DDACTS[5].

It is important to realize that even if there is no causal link between crimes and crashes the benefits of enforcement on one can affect the other due to the presence of the police to be visible and/or taking action on one or the other.

There are several data sets available to the analyst to conduct this type of analysis. Many of these data sets are available through other administrative services of the jurisdiction.

- Parcel
- Tax Assessor
- Land Use

[5] Harris P., et. al. (2009). *Investigating the Simultaneous Effects of Individual, Program and Neighborhood Attributes on Juvenile Recidivism Using GIS and Spatial Data Mining.* Final Report to the National Institute of Justice, Grant #2006-IJ-CX-0022.

- Zoning
- Code Enforcement
- Building Footprints

A general method:

1. Place correlation test for accidents and crime types, accidents and violation types, and crime and violation types.
2. Cluster tests individually.
3. Cluster tests simultaneously between accidents and crime types, accidents and violation types, and crime and violation types.
4. Density surface modeling individually.
5. Density surface modeling simultaneously between accidents and crime types, accidents and violation types, and crime and violation types.
6. Examine locations where the modeled paths converge and their characteristics. This would include an overlay of the modeled paths and cluster analysis.
7. Compare results from each hot spot through descriptive statistics.
8. Examine the characteristics of where offenders come from for violations, accidents and crimes.
9. Characterize places qualitatively.
10. Conduct near repeat analysis on each hot spot individually.
11. Need to look at the criminal and traffic violation histories of individuals, which includes previous accidents and where.
12. Model travel paths of violations, accidents and crime trips.