Conjunctive Analysis Report: 2012 Motor Vehicle Theft in Colorado Springs

PIs Leslie Kennedy, Joel Caplan & Eric Piza

Completed in partial fulfillment of National Institute of Justice (NIJ) award #2012-IJ-CX-0038

Background and Overview

Research suggests that crime is not evenly distributed throughout the environment (Sherman, Gartin, and Buerger, 1989). This can be explained through the presence of *attractors* and *generators* of crime (Brantingham and Brantingham, 1995), or "risk factors," that co-locate to create unique "behavior settings" (Taylor, 1997) that are conducive to illegal activities. Risk terrain modeling (RTM) is an approach to risk assessment that uses existing technology, data, and GIS (geographic information systems) to diagnose the underlying characteristics of the environment that produce crime-conducive behavior settings. RTM standardizes risk factors to common geographic units over a continuous surface and combines multiple risk factors to produce a composite map showing the presence or absence of risk at micro-level places (Caplan, Kennedy, Miller, 2010). It is used to empirically test ideas about emerging conditions leading to crime problems, to develop interventions that deploy police officers to high risk places, and to help prioritize crime risk factors for mitigation efforts. However, the conditions that influence crime are highly complex.

It is likely that some interaction effects among certain risk factors in a RTM are stronger than other interactions on the attraction of criminal behavior, even for weighted risk terrain models. Missing from the current outputs of RTM methods and statistical validation tests is an easy way to determine which combination of risk factors account for the greatest relative frequency of crimes compared to all other risk factor combinations. For example, regression modeling (i.e., to test predictive validity) provides information about the likelihood of crime occurring at places with every increased risk value. If we can imagine an un-weighted 4-factor RTM, places with all 4 risk factors present will have the greatest risk, or likelihood, of crime occurring there. However, places with risk values of "4" should not be treated equal because it is unclear which factors' absence makes the "best" 4-factor model. Should places with factors A, B, C, D be prioritized over places with factors B, C, D, E or vice versa? In this example, several places can have risk values of "4" but have meaningfully different combinations of risk factors. We need a way to test which behavior settings (e.g., bars + parks + schools *or* bars + parks + fast food restaurants) account for the most crime events.

Conjunctive analysis can be used to explore the relative interaction among risk factors by allowing for the comparison of cross-case configurations and by providing empirical evidence demonstrating the interrelationships between factors of the location in question. It provides a multivariate analysis of discrete categorical data that can be used for various criminal justice applications, such as to examine the risks of imprisonment for federal drug offenders (Miethe, Hart, and Regoeczi, 2008). The end product or output of a conjunctive analysis is a data matrix of behavior settings, which include every possible combination of risk factor interactions and the relative frequency of crimes associated with each setting. Conjunctive analysis can be used to enhance the practical value of RTM by addressing the interrelationships among different factors. More specifically, RTM can be used to empirically identify and validate environmental risk factors. These significant risk factors can then be incorporated into a conjunctive analysis to assess interaction affects among them.

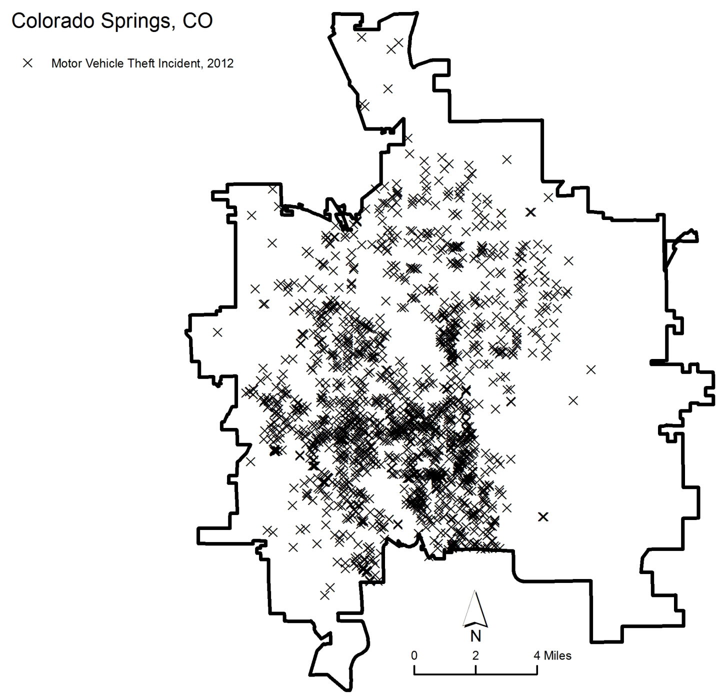
The purpose of this report is to combine the power of RTM and conjunctive analysis to examine the spatial context of motor vehicle theft in Colorado Springs, CO. This report will first describe the steps involved in building a RTM for motor vehicle theft in Colorado Springs, as well as the results of this model. Subsequent sections will present the conjunctive analysis process, along with the results of a conjunctive analysis that was informed by results from the RTM. Finally, the practical implications of using conjunctive analysis and RTM to inform targeted police interventions are discussed.

Building the Risk Terrain Model

As discussed above, RTM identifies the risky features of a landscape and models how they help to create unique behavior settings for crime. The Risk Terrain Modeling Diagnostics (RTMDx) Utility was used to produce a risk terrain model[1] for the 1,959 incidents of motor vehicle theft that occurred in calendar year 2012. The RTMDx Utility is a software app produced by Rutgers University that automates the RTM process[2].

Figure 1 is a pin map displaying the distribution of motor vehicle theft incidents in the jurisdiction of Colorado Springs, CO. It appears that incident locations of this crime do not distribute evenly throughout the jurisdiction. As depicted on the map, crimes seem to cluster at certain areas. By diagnosing the underlying spatial factors of motor vehicle theft at existing high-crime places, RTM helps to forecast where crime is statistically most likely to occur in the future and provides actionable information for focused interventions. The following sections will discuss the basic steps of building the RTM used for this conjunctive analysis, including selecting likely risk factors, setting parameters in the RTMDx Utility, and operationalization of the risk factors. The results of the risk terrain model are then presented.

Figure 1. 2012 Incident Locations of Motor Vehicle Theft in Colorado Springs, CO



Selecting Likely Risk Factors

With any RTM analysis, the product of this task should be a comprehensive list of risk factors related to the outcome event. The RTMDx Utility can accept up to 30 risk factors as inputs for testing. Three methods were used to identify the final "pool" of risk factors that should be tested for inclusion in the RTM. 1) Existing empirical research literature was reviewed, which identified a variety of risk factors that have been found to correlate with motor vehicle theft. 2) Professional/practitioner insights also played an invaluable role in determining which factors are likely relevant for this particular jurisdiction. Jurisdictions may have specific crime problems that relate to a unique set of factors that are best identified by local officials who work in these environments on a regular basis.

Once a pool of potential risk factors were compiled, 3) visual inspection in a GIS was used to explore which features appear to co-locate with crime incident locations. Using ArcGIS, each potential risk factor data set was layered, respectively, with point features of motor vehicle theft incidents and visually inspected for spatial relationships. Risk factors that generally appeared to share a spatial relationship were retained for empirical testing in the RTMDx Utility. This process was used to exclude risk factors that very obviously did not visually appear to spatially relate to the crime incidents, in order to create a more parsimonious model.

Risk Factors and Data Sources

Risk terrain modeling relies on valid and reliable data sources. Data used for this study was collected from InfoGroup and the Colorado Springs Police Department (CSPD). InfoGroup is a data and marketing services company that provides detailed information about public entities. Some of the risk factor data sets used in this case study were obtained from InfoGroup at the address-level: bowling centers, hotels and motels, night clubs, parking stations and garages, sit down restaurants, take out restaurants, retail shops, malls, foreclosures, gas stations with convenience stores, liquor stores, variety stores, and convenience stores. Crime and other data sets were provided at the address-level by the CSPD: motor vehicle theft incidents, bars, parks, schools, disorder calls for service, multifamily housing, and commercial zones.

Setting Parameters in the RTMDx Utility

Several parameters must be set in the RTMDx Utility before testing can begin. For this analysis, an "aggravating" model type was run to determine the underlying attractors of motor vehicle theft in Colorado Springs, CO. Other parameters include "block length," "cell size," "maximum spatial influence," "analysis increments," and "operationalization," which are discussed below.

Block Length, Cell Size, Maximum Spatial Influence, and Analysis Increments

Risk terrain modeling is typically concerned with the micro-level unit of analysis, such as a raster GRID cell. In this case, "block length" was set to the mean length of a block face in Colorado Springs (551 feet) and the "cell size" was set to half of the block length (276 feet). Previous empirical research suggests that the spatial influence of a given environmental feature extends no more than just a few street blocks. Therefore, the "maximum spatial influence" for this model was set to 3 blocks. The RTMDx Utility allows for testing either whole or half block increments. Each risk factor in this study was tested at whole block increments.

Operationalization

The "operationalization" parameter was custom selected for each risk factor tested. This parameter regards how each risk factor's spatial influence will be assessed; that is, as a function of "proximity," "density," or "both." "Proximity" proposes that being within a certain distance of the factor features increases the likelihood of crime. "Density" proposes that risk is higher at places where the factor features are heavily concentrated. The RTMDx Utility also allows for testing "both" proximity and density. This setting permits the Utility to empirically select the best operationalization. However, this doubles the number of variables to be tested and greatly increases the analytical run time. So, to purposefully make this parameter decision for each risk factor, output results from the Nearest Neighbor (NN) analysis tool in ArcGIS were consulted. Table 1 shows the results of the NN analysis and the corresponding "operationalization" parameter selected for each risk factor.

Table 1. Results of NN Analysis and Corresponding Risk Factor Operationalizations

Name	Operationalization	Observed Distance	p-value	Spatial Pattern
Bars	Both	2112.28	0.00	Clustered
Parks	Proximity	159.68	0.00	Clustered
Schools	Proximity	2890.99	0.34	Random
Bowling Centers	Proximity	12456.37	0.01	Dispersed
Hotels and Motels	Both	1332.10	0.00	Clustered
Night Clubs	Proximity	6083.29	0.07	Dispersed
Parking Stations and Garages	Proximity	14738.25	0.00	Dispersed
Sit Down Restaurants	Both	518.23	0.00	Clustered
Take Out Restaurants	Both	695.55	0.00	Clustered
Retail Shops	Both	2278.92	0.00	Clustered
Malls	Proximity	10210.22	0.03	Dispersed
Foreclosures	Both	764.36	0.00	Clustered
Gas Stations w/ Conv. Stores	Proximity	6189.57	0.06	Dispersed
Liquor Stores	Proximity	3016.24	0.07	Clustered
Variety Stores	Proximity	5390.24	0.04	Dispersed
Disorder CFS	Density	62.30	0.00	Clustered
Multifamily Housing	Proximity	146.38	0.00	Clustered
Commercial Zoning	Proximity	224.08	0.00	Clustered
Convenience Stores	Proximity	3281.73	0.96	Random

To Produce Table 1:
First, a "nearest neighbor threshold" was calculated as: *NN Threshold = 2 * (Block Length * Number of Analysis Increments).* The NN threshold for this case study is 3,306. Then, for each set of risk factor feature points: If the features were not significantly clustered ($p>0.05$) or if the observed mean distance reported by the Nearest Neighbor analysis was greater than the NN threshold, the operationalization was set to "proximity." If the points were significantly clustered and the observed mean distance was less than or equal to the NN threshold, "both" proximity and density was used, allowing the model to empirically determine the best one, if any. There were, however, some exceptions to this general rule. Risk factor data sets that represented a fleeting phenomenon (i.e., they occurred at a location, but did not remain a permanent feature of the environment), such as drug arrests or calls for service, were tested as a function of "density." And, because the RTMDx Utility supports only point features as inputs, some polygon shapefiles had to be converted to (representative) point features prior to being tested (e.g., parks were typically polygons shapefiles and were converted to point features). Given the method of conversion, these risk factors were tested as "proximity" only.

Results of the Risk Terrain Model
There were 77,873 raster GRID cells (i.e., 276ft x 276ft) used in the analysis, 1,465 of which contained crime incidents. The RTMDx Utility identified 7 statistically significant risk factors for motor vehicle theft incidents in Colorado Springs, CO and produced a risk terrain map (Figure 2) with relative risk values at each micro-level place. For a more detailed explanation of the statistical procedures see the *RTMDx Utility User Manual*. Relative risk values (RRVs) ranged from 1 (for the lowest risk place) to 163 (for the highest risk place). The highest risk places are 163 times more likely to experience motor vehicle theft than the lowest risk places.

Locations that are two standard deviations above the mean RRV (i.e. places shaded black on the map) are considered the highest risk locations for motor vehicle theft in Colorado Springs, CO. The likelihood of crimes occurring at these places is more than 49 times higher than some other locations.

Figure 2. RTM for Motor Vehicle Theft in Colorado Springs, 2012

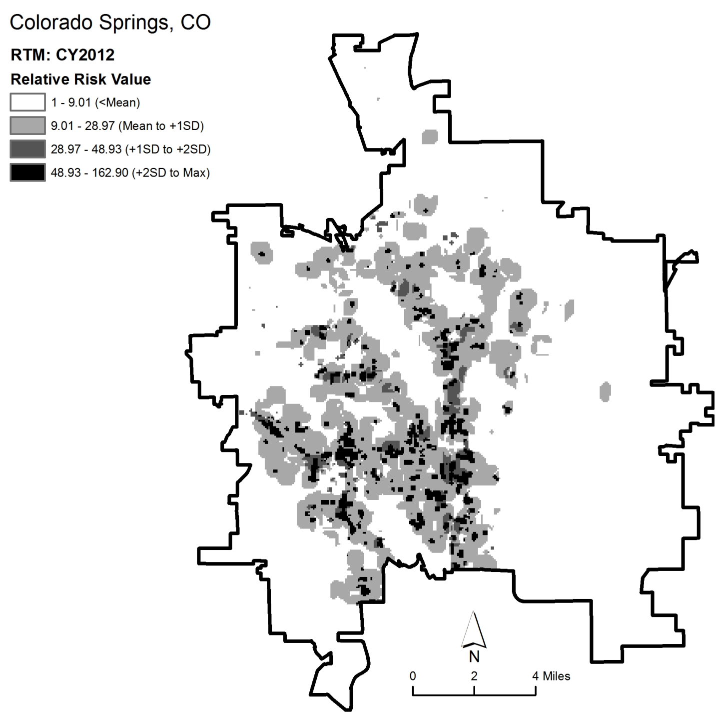
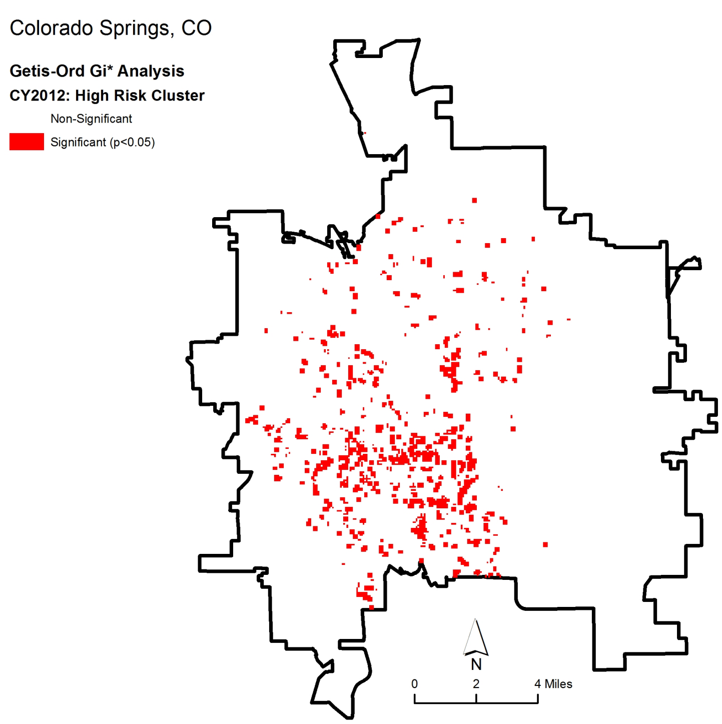


Figure 3. Getis-Ord Gi* Analysis of High Risk Clusters for Colorado Springs, 2012



The Hotspot Analysis tool (Getis-Ord Gi*) in ArcGIS was used to identify statistically significant clusters of relative risk values (i.e., high-risk clusters). As Figure 3 illustrates, there are significant clusters of high-risk in Colorado Springs, CO.

The 7 risk factors included in the risk terrain model are disorder calls for service, multifamily housing, foreclosures, parks, sit down restaurants, commercial zoning, and convenience stores. The most meaningful operationalizations and spatial influential distances of these factors are presented in Table 2. The relative risk values can be interpreted as the weights of risk factors and may be easily compared. For instance, a place influenced by disorder calls for service has an expected rate of crime that is more than 4 times higher than a place influenced by convenience stores (RRVs: 5.31 / 1.27 = 4.18). The most important predictor of motor vehicle theft occurrence is dense locations of disorder calls for service. Accordingly, all places may pose a risk of motor vehicle theft to people in Colorado Springs, CO, but because of the spatial influence of certain features of the landscape, some places are riskier than others.

Table 2. Risk Terrain Model Specifications

Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Disorder CFS	Density	551	1.67	5.32
Multifamily Housing	Proximity	1653	1.00	2.72
Foreclosures	Proximity	1653	0.96	2.60
Parks	Proximity	1653	0.58	1.78
Sit Down Restaurants	Proximity	1653	0.34	1.41
Commercial Zoning	Proximity	1653	0.31	1.37
Convenience Stores	Proximity	1653	0.24	1.27
Intercept	--	--	-5.91	--
Intercept	--	--	-0.57	--

Conjunctive Analysis

Once a significant risk terrain model is identified, conjunctive analysis can begin. Conjunctive analysis is a fairly straightforward statistical technique. According to Miethe et al. (2008),

> "A conjunctive analysis of case configurations begins with an aggregated compilation of all possible combinations of attributes considered simultaneously. The number of possible case configurations depends on the number of independent variables and categories within them. For a conjunctive analysis involving 5 dichotomous independent variables, there are 32 qualitatively distinct case configurations (2^5=32)…Once the possible case configurations are identified, conjunctive analysis proceeds by aggregating each observation into their respective case configuration and exploring the relative distribution of particular categories of the outcome variables across these configurations" (p.229).

The data for the conjunctive analysis was prepared using ArcGIS. In this step, the spatial influence of each significant risk factor was coded as a dichotomous variable representing the presence (1) or absence (0) of highest risk at each micro-place (i.e., 276x276 raster cell) in the study area. Given 7 binary independent variables, the total number of possible case configurations for this conjunctive analysis is 128 (2^7=128). The final data was imported into SPSS (Statistical Packages for the Social Sciences) to perform a conjunctive analysis using the following formula (Miethe et al., 2008):

```
AGGREGATE
/OUTFILE = 'CA_Matrix_file'
/BREAK = A B C D
/Crime = SUM
/N_Cases = N
```

The resulting product is a conjunctive analysis data matrix (Miethe et al., 2008) that displays all of the possible case configurations of the aggregated compilation of risk factors. When displayed in a table of *i* rows and *j* columns, each row represents a particular case configuration. Each row also includes the number of observations (i.e., count of 276ft x 276ft cells) and the proportional distribution of outcome events for that unique case configuration. As explained by Miethe et al. (2008), "conjunctive analysis involves visual representations of case configurations that convey important information about their nature, diversity, and distribution for subsequent analysis" (p.229).

Conjunctive Analysis Results

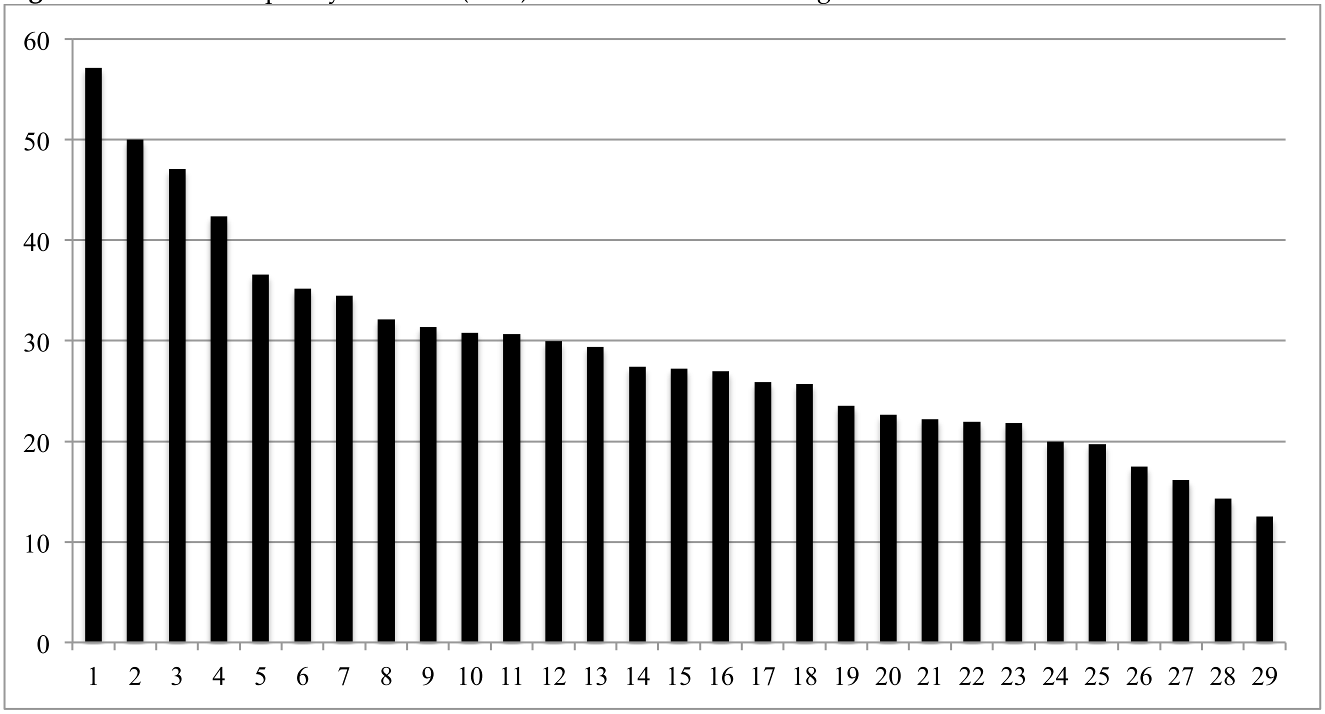
A conjunctive analysis data matrix for the 7 significant risk factors for motor vehicle theft in Colorado Springs, CO is displayed in Table 3. Of the 128 possible case configurations, a total of 115 were observed. However, only dominant case configurations (>9 observations) (Miethe et al., 2008, p.229) with a relative frequency of crime (RFC) above the mean (11.43) are displayed. This results in a total of 29 case configurations that are of particular interest to us. Each row under the case configuration column in Table 3 refers to a behavior setting defined by a unique set of attributes (i.e., presence or absence of risk factors). Each risk factor has a column containing a series of 1s and 0s indicating the presence or absence of that risk factor's spatial influence in each case configuration. For example, case configuration 1 is characterized by *the presence of* the spatial influences of disorder calls for service, multifamily housing, foreclosures, sit down restaurants, and *the absence of* the spatial influences of parks, commercial zoning, and convenience stores.

Also from Table 3, we see that there are 14 observed instances (i.e., cells) of configuration 1, which were responsible for 8 of the 1,959, or 0.41%, of motor vehicle thefts in 2012. The RFC of case configuration 1 is 57.14. This represents the most influential of all case configurations displayed in Table 3. Finally, Table 3 includes a break to denote case configurations with a RFC that is one or more standard deviations above the mean RFC. The 13 case configurations with a RFC greater than one standard deviation above the mean were responsible for 521 of 1,959 motor vehicle thefts, or a total of 26.60%. Of the 77,873 raster cells analyzed, these 13 case configurations included 1,575 cells, or 2.02% of the total study area. The 29 case configurations with a RFC that was higher than the mean accounted for 845 of 1,959 motor vehicle thefts, or 43.13%. These 29 case configurations included 2,953 cells, or 3.79% of the total study area. Thus, behavior settings covering less than 4% of the study area account for nearly 43% of all crime incident locations.

Table 3. Conjunctive Analysis Data Matrix of Dominant Case Configurations with a Relative Frequency of Crime Above the Mean

Case Configuration	Disorder CFS	Multifamily Housing	Foreclosures	Parks	Sit Down Restaurants	Commercial Zoning	Convenience Stores	Crime Count (total n=1,959)	# of Case Configurations	% Crime	Relative Frequency of Crime (RFC)
1	1	1	1	0	1	0	0	8	14	0.41	57.14
2	1	0	0	1	0	1	0	7	14	0.36	50.00
3	1	1	1	0	0	1	1	8	17	0.41	47.06
4	1	1	0	0	1	1	1	25	59	1.28	42.37
5	1	1	1	1	0	1	1	15	41	0.77	36.59
6	1	1	1	1	1	1	1	146	415	7.45	35.18
7	1	0	0	0	1	1	0	10	29	0.51	34.48
8	1	1	0	1	0	1	0	9	28	0.46	32.14
9	1	1	1	1	1	1	0	116	370	5.92	31.35
10	1	0	0	0	1	1	1	4	13	0.20	30.77
11	1	1	1	1	0	1	0	46	150	2.35	30.67
12	1	1	1	1	1	0	1	112	374	5.72	29.95
13	1	1	1	1	0	0	1	15	51	0.77	29.41
One Standard Deviation								**521**	**1575**	**26.60**	
14	1	0	1	1	1	0	1	20	73	1.02	27.40
15	1	1	1	1	0	0	0	49	180	2.50	27.22
16	1	1	1	0	1	0	1	24	89	1.23	26.97
17	1	1	0	1	1	1	0	30	116	1.53	25.86
18	1	1	1	0	1	1	0	46	179	2.35	25.70
19	1	0	1	0	1	1	0	4	17	0.20	23.53
20	1	1	1	1	1	0	0	53	234	2.71	22.65
21	1	1	0	0	1	1	0	6	27	0.31	22.22
22	1	1	0	1	1	0	0	9	41	0.46	21.95
23	1	1	1	0	1	1	1	46	211	2.35	21.80
24	1	1	0	1	0	0	0	4	20	0.20	20.00
25	1	1	0	1	1	1	1	15	76	0.77	19.74
26	1	0	1	0	1	1	1	7	40	0.36	17.50
27	1	0	1	1	1	1	1	5	31	0.26	16.13
28	1	0	1	1	0	0	0	4	28	0.20	14.29
29	1	1	1	0	0	1	0	2	16	0.10	12.50
Mean								**845**	**2953**	**43.13**	

Figure 4. Relative Frequency of Crime (RFC) for each Behavior Setting



Discussion

Conjunctive analysis can be used to explore which unique combinations of risk factors identified via risk terrain modeling attract the most incidents of crime. Of the 128 possible configurations, we highlight 29 dominant case configurations with a RFC that is greater than the mean RFC. The most influential behavior setting (i.e., case configuration 1) is marked by the presence of four risk factors: disorder calls for service, multifamily housing, foreclosures, sit down restaurants. When the spatial influential areas of these factors exist together, these behavior settings account for 0.41% of motor vehicle thefts in Colorado Springs, CO. While this seems like a small portion of motor vehicle theft, it is important to consider the RFC. For example, case configuration 55 (not displayed) accounts for the largest raw portion (7.96%) of motor vehicle theft in Colorado Springs, CO. However, RFC takes into account the amount of crime in relation to the relative size of the geography involved in its occurrence. Thus, case configuration 1 exerts nearly 17 times the spatial influence of case configuration 55 (57.14 / 3.34 = 17.11). The matrix produced by the conjunctive analysis provides a visual tool that highlights how risk factors co-locate at micro-level places to create unique behavior settings that are attractive to criminal behavior. These behavior settings can then be mapped in a GIS for analytic purposes or resource allocation.

Risk terrain modeling identified the 7 risk factors (disorder calls for service, multifamily housing, foreclosures, parks, sit down restaurants, commercial zoning, and convenience stores) that pose the highest level of risk for motor vehicle theft in Colorado Springs, CO. Conjunctive analysis identified places with particular subsets of these 7 factors (e.g., A, B, C, D) to be prioritized over places with *different* and less risky combinations of these factors (e.g., B, C, D, E). Several places can have risk values of, for example, "4" but have meaningfully different combinations of risk factors. For example, case configurations 1 and 10 each have a set of four risk factors. However, case configuration 1 (disorder calls for service, multifamily housing, foreclosures, and sit down restaurants) is qualitatively different from case configuration 10 (disorder calls for service, sit down restaurants, commercial zoning, and convenience stores). Using information obtained from the conjunctive analysis data matrix, we can see that case configuration 1 has nearly 1.86 times the relative frequency of crime as case configuration 10. From a practical perspective, conjunctive analysis verifies the important risk factor interactions, allowing police to make better decisions regarding the most effective allocation of resources and intervention strategies at specific behavior settings.

Risk terrain modeling provides public safety practitioners with a tool to empirically identify which features of the environment influence illegal behavior. Moreover, it allows agencies to develop targeted, risk-based strategies according to the weights given to each significant factor in the model. However, the effectiveness of police interventions will likely vary as a function of the particular behavior setting that is being targeted because the interaction of some risky features from RTM are likely to be more prone to crime than the interaction of other risky features from the same model. Conjunctive analysis can be used to evaluate the un-weighted interactions among the risk factors within a RTM to determine which combination of features constitute the most influential behavior settings to be sure that resources are geographically targeted in the most efficient manner. Future research should evaluate the use of conjunctive analysis as a way for police to assess their long-term effectiveness in reducing crime. This can be done by conducting several conjunctive analyses over a length of time to determine which intervention activities were most effective at which behavior settings. Interventions can be updated and redeployed based on information obtained from each conjunctive analysis. Using conjunctive analysis in this way, police can continually ensure that they are employing the most effective strategies in the most appropriate places.

References

Brantingham, P., & Brantingham, P. (1995). Criminality of place. *European Journal on Criminal Policy and Research*, *3*(3), 5-26.

Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. *Justice Quarterly*, *28*(2), 360-381.

Miethe, T. D., Hart, T. C., & Regoeczi, W. C. (2008). The conjunctive analysis of case configurations: An exploratory method for discrete multivariate analyses of crime data. *Journal of Quantitative Criminology*, *24*(2), 227-241.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, *27*(1), 27-56.

Taylor, R. B. (1997). Social order and disorder of street-blocks and neighborhood: Ecology, microecology and the systemic model of social disorganization. Journal of Research in Crime and Delinquency, 24, 113–155.

Taylor, R. B., & Harrell, A. (1996). Physical environment and crime (pp. 11-12). US Department of Justice, Office of Justice Programs, National Institute of Justice. Retrieved Mar 26 2014, From: https://www.ncjrs.gov/pdffiles/physenv.pdf

Endnotes

[1] The risk terrain model used for this conjunctive analysis was updated after the NIJ project intervention in the city had begun because there were recent improvements to the modeling process and new access to better data sets. So, the risk terrain model presented in this conjunctive analysis report might differ slightly from the model that originally informed the intervention in this city.

[2] http://www.rutgerscps.org/software/index.html

