Fall 2011

Investigating Factors Associated with Burglary Crime Analysis using Logistic Regression Modeling

Daniel Antolos
Embry-Riddle Aeronautical University - Daytona Beach

Investigating Factors Associated with Burglary Crime Analysis using Logistic Regression Modeling

by

Daniel Antolos

B.S. Embry-Riddle Aeronautical University, 2002

M.A. Webster University, 2007

A Graduate Thesis Submitted to the

Department of Human Factors and Systems

in Partial Fulfillment of the Requirement for the Degree of

Master of Science in Human Factors and Systems

Embry-Riddle Aeronautical University

Daytona Beach, Florida

Fall 2011

Investigating Factors Associated with Burglary Crime Analysis using Logistic Regression Modeling

by

Major Daniel Antolos

This thesis was prepared under the direction of the candidate's thesis committee chair, Dahai Liu, Ph.D., Department of Human Factors and Systems, and has been approved by the members of the thesis committee. It was submitted to the Department of Human Factors and Systems and has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Human Factors and Systems.

THESIS COMMITTEE:

Dahai Liu, Ph.D., Chair

Shawn Doherty, Ph.D., Member

Andrei Ludu, Ph.D., Member

MSHFS Program Coordinator

Department Chair, Department of Human Factors and Systems

Associate Vice President for Academics

1-19-2012

Acknowledgements

The author wishes to express his extreme gratitude to his thesis chair Dr. Dahai Liu. Without his help, guidance and patience this study would not have been possible.

Additionally, thanks are in order to Dr. Shawn Doherty and Dr. Andrei Ludu for their continued efforts in this endeavor. They have both given their time to make a considerable impact on this study through their guidance and input.

The author would like to acknowledge the professors of the Human Factors and Systems department, their enthusiasm and passion for their profession was a motivator to return to Embry-Riddle Aeronautical University to pursue a Master of Science in Human Factors and Systems as well as expand upon his professional knowledge as a United States Army officer.

Finally, the author also wishes to thank his wife, Allison Antolos. She has consistently provided support and encouragement in all his endeavors in their time together. Aside from being incredibly supportive of everything, she has remained understanding and unwavering in the trials and tribulations they have faced together.

Table of Contents

List of Figures

List of Tables

Abstract

Author: Daniel Anthony Antolos

Title: Investigating Factors Associated with Burglary Crime Analysis using Logistic Regression Modeling

Institution: Embry-Riddle Aeronautical University

Year: 2011

This study conducted a logistic regression to determine the relationship of factors associated with burglary to determine the variables necessary to predict criminal activity. Predictors utilized in the study; included time of day, day of week, connectors, barriers, and repeat victimization. These predictors were all incorporated to develop a model that would best predict burglary activity as it relates from a determine epicenter of activity. The predictors selected have all be shown, through research to be significant, characteristics of activity as they relate to burglary but have not been incorporated together to develop a significant model. The model compared the strength of the predictors as they relate to the distance they occur from the identified epicenter of a density plot. Data was collected from a statistically significant area associated with burglary activity. Additional information was extrapolated from criminal activity records provided by the local law enforcement. The predictors and model were statistically evaluated to determine the significance of the predictors and the ability of the model to accurately predict burglary activity within the sampled area. Analysis showed the model was significant, however the comparison in area size in association with the predictors showed to be insignificant. Multiple comparisons of area sizes were observed only to discover that greater comparison in area yielded less significant results. Further research would benefit by observing smaller clusters of activity within a one kilometer area utilizing the same predictors.

Introduction

Crime Analysis

History of crime analysis.

Informal crime analysis, in its simplest sense, is performed by all officers when a crime is being investigated. Crimes analysis is the value of examining a crime's occurrence and comparing it with similar past events with the intent of identifying similar factors with the intent of establishing a strategy to combat the criminal behavior. All law enforcement professionals are trained in the same way; the lone police officer is a walking crime analysis unit as they compare their investigations with his past experiences and with the experiences of others (Buck, 1973).

Crime analysis started in 1847 with the London Metropolitan Police when they determined that certain characteristics of crime and the motives of the persons who committed them developed some form of pattern. Modus Operandi or Method of Operation was coined to classify offenders and the crimes they committed (A History of Crime Analysis, 2011).

August Vollmer, who is considered the father of American policing, developed new innovations which became the basis for modern policing. In 1906 he instituted a basic records management system which organized police reports based on type of crime committed. To ensure these reports and records were not lost to time, he mandated a regular review of the reports as well as mapping the locations of each crime on a map with a distinctive color pin. Based on the picture that was developed he was able to establish patrol districts based on certain types of crimes and the details associated with that crime (A History of Crime, 2011).

Vollmer stated that crime analyst must observe the new crimes committed and compare them with old occurrences (A History of Crime, 2011). He believed a pattern or trend would emerge that would provide police with an added advantage in preventing and solving crime more efficiently.

Vollmer determined he needed to develop a classification and categorization method for the criminal report system he developed. Early attempts were inefficient and were based on statute violations, which in the early 1900's could be conflicted based on different jurisdiction classification of counties, cities, and states (A History of Crime Analysis, 2011).

The Federal Bureau of Investigation (FBI) agreed with Vollmer's classification efforts and developed a more efficient means of categorizing the information. In 1922 the Uniform Crime Reporting Program was established which classified individual statutes into more broad and more incumbent categories (A History of Crime Analysis, 2011).

Over the years following the establishment of the Uniform Crime Reporting Program, crime analysis became a driving force in conducting police work throughout the United States. Orlando Wilson, a protégé of Vollmer, saw fit to develop a book in 1963, *Book on Police Administration*, that would be utilized as a guide and standardize crime analysis throughout the United States:

> "The crime-analysis section studies daily reports of serious crimes in order to determine the location, time, special characteristics, similarities to other criminal attacks, and various significant facts that may help to identify either a criminal or the existence of a pattern of criminal activity. Such information is helpful in the planning the operations of a division or district."

Increase in crime.

Vollmer and Wilson's efforts were implemented and utilized for many decades with great success. Their knowledge, expertise, and implementation of their practices helped to develop the U.S. law enforcement capabilities, making them more successful. However, in the 1960s and 1970s there was an apparent rise in crime nationwide. Law enforcement agencies and congress noticed there was a shortfall of funding and capability due to the lack of funding to combat the rise in the crime statistics throughout the U.S.

In 1965 the Law Enforcement Assistance Act (LEAA) was instituted and provided funding for the establishment of crime analysis units and the resources needed to make those units successful (A History of Crime Analysis, 2011). Under this new act the National Advisory Commission on Criminal Justice Standards and Goals recommended crime analysis capability was needed in every police agency. In 1970 disparate crime reduction and crime prevention programs became united. This integrated the Criminal Apprehension Program (CAP) in such areas as crime, intelligence, investigation, operations analysis which was similar to established programs in California and Massachusetts.

As the LEAA was dissolved, support was shifted to the Office of Justice Programs in 1984 which continued the same type of analysis as originally instituted by the LEAA.

Crime analysis today.

In the 1990's crime analysis took a new direction based on the book *Problem Oriented Policing* by Herman Goldstein. Goldstein was an executive assistant to Wilson when he was the chief of the Chicago Police Department in the 1960s. Due to his expertise and knowledge developed while working for Wilson, he was able to develop a handbook that is still widely used by law enforcement professionals today. He established that the focus should not be on the crime incident but more on the crime problem. The law enforcement profession needs to eliminate root causes before the real crime problem develops. It was the job of the crime analyst to identify and analyze these crime problems and determine those root causes (Goldstein, 1990).

Technology for crime analysis.

With the assimilation of this Goldstein's handbook, new technology became available for crime analysts to utilize and assist them with Goldstein's practices. Programs such as word processors, database management, communication, and presentations assisted analyst to develop a common operating picture of the crime problem that can be distributed and analyzed by all law enforcement professionals.

The biggest leap to date in crime analysis has been the institution of crime mapping, more specifically Geographic Information Systems (GIS). GIS provided capabilities such as raster mapping; utilizing topographic survey software to map out changes in crime densities in a given area. Buffer analysis allowed analysts to generate a buffer around targeted areas and then identifying or select features based on whether they fall inside or outside the boundary of the buffer (A History of Crime, 2011). Three dimensional imaging allowed law enforcement professionals to capture specific details of a crime scene immediately for further analysis after the scene had been processed. This allowed law enforcement professionals to revisit the scene in the same condition it was when discovered.

In 1994 the New York Police Department incorporated a GIS system named the "CompStat" or comparative statistics. This system allowed the NYPD to conduct extensive crime analysis as well as strategy development, and management accountability utilizing crime mapping data (A History of Crime Analysis, 2011).

In 1997 the National Institute of Justice Crime Mapping Research Center started the promotion, research, evaluation, development, and dissemination of GIS technology and spatial analysis of crime data to provide the U.S. law enforcement community with tools necessary to reduce crime.

Criminal Behavior

Inherent criminality.

Sterzer (2010) attempted to determine if criminality is an inherent property or programmed into a person upon their birth. Their research sought to identify potential biological factors in the etiology of criminal behavior. Sterzer was not able to determine if there were actual biological markers to predispose criminality but found there was more of a conditioning aspect to the development to develop antisocial behavior. Poor fear conditioning, or not being exposed to situations exercising the ability to understand the consequences associated with poor decision making, was one of the leading contributors to antisocial behavior showed that due to poor ethical conditioning, fearless individuals had deficient amygdale: the part of the brain that is associated with fear.

Psychology of criminal behavior.

Criminology is a psychological perspective that emphasizes the situational specificity of behavior. Mischel and Shoda (1995) showed that how a person approaches the world is not predetermined by traits that will cause a person to act a certain way. It is more how the person interacts with their environment; the manner in which they conduct themselves and the results of their behavior in regards to the situation's outcome.

Self control can be one of those traits that are developed through environment interaction that can have a profound effect upon a person and their development. *A General Theory of Crime* (Gottfredson & Hirschi, 1990) states that self control is a dominant trait-like characteristic that is considered when discerning criminal behavior. Self control is one of those traits that are shaped by context as a natural product of learning. More specifically, it can be determined that self control is seen as stimulus control in a complex learning environment (Horney, 2006).

Observing clusters of specific types of situations can provide an insight to how a person will develop and ultimately react to future situations. In this case looking for environmental consistencies in an individual's life will provide evidence to how they conduct themselves in given and future situations (Horney, 2006).

Criminality is seen as the underlying behavior of an individual to commit some form of crime based on motivation. Crime is the result of that behavior when situations produce themselves and allow the individual to commit those crimes (Horney, 2006).

Laub (2005) showed that criminal behavior is developed over time through multiple situations and the outcomes of those situations. Only then the person develops a mindset of when and how to commit crimes.

Human development and crime.

Juveniles reported that peer influence and peer pressure, provocation and anger, boredom and thrill, alcohol and drugs, and money were the motivators for committed offenses (Kai, Heng, & Bullock, 2007). Research has also shown that an apparent lack of success in school can be one of the principle causes of criminal behavior (Trout, Nordness, Pierce & Epstein, 2003; Zabel & Nigro, 2001). Without success in an academic institution and the influences of peers and materialism a person will perceive being shortened by opportunities to progress in life. This does not promote a positive mental health environment and can be further negatively influenced by other motivators such as drugs and alcohol.

Deficit theory, poor reception of social cues, lack of impulse control and lack of the ability to learn from experience increase an individual's susceptibility to delinquent behaviors (Kai et al., 2007). Such things as lack of supervision and poor or no positive role models can contribute to negative behavior development. Motivations such as earning quick money for either necessities or amenities can contribute to criminal behavior.

Complex forms of human functioning are learned and that aberrant behavior is related to the context in which it occurs when discussing complex human processing (Alberto & Troutman, 2003). In youth this aberrant behavior is learned from consequences from behavior demonstrated in past occurrences (Kai et al., 2007).

In the Kai et al. (2007) study, certain behavior was observed to determine the motivations to commit specific types of crime. Peer influence and peer pressure were prevalent in crimes associated with violence involving individuals and gangs. Provocation and anger were involved in crimes were the person was not able to control their rage or harassment towards others. Boredom and thrill were seen as motivators for violent crimes but more so in nonviolent crimes such as burglary and theft.

Trejos-Castillo, Vazsonyi, and Jenkins (2008) showed that self-esteem, risk-proneness, and educational commitment in combination will have a more profound effect of a person being involved in violent and criminal behavior. Low self-esteem is a contributor to aggression, delinquency, and violence. Risk-proneness is the attraction to excitement and risk and a lack of awareness of negative consequences which is also associated with aggression. Hawkins, Guo, Hill Battin-Pearson, and Abbott (2001) and Swaim, Kimberly, & Kathleen (2006) showed that school connectedness, academic performance, educational commitment, and career expectations have been consistently cited as key predictors of problem behaviors.

Crime as it relates to the brain.

Krandel and Freed (1989) demonstrated that frontal lobe dysfunction is a contributor towards criminal behavior, but could not be associated with any particular form of crime. Brower & Price (2001) were able to determine that 56.9 percent of habitually aggressive subjects had EEG abnormalities compared with 11.8 percent of other subjects who had committed a single, isolated aggressive act. Clinical signs of frontal lobe dysfunction are prevalent in populations of persons prone to violent and antisocial behavior. This supports other studies which reported antisocial personality traits after frontal lobe injury, more specifically the orbital-frontal cortex.

Criminal decision making.

As discussed in regards to motivators of criminal behavior in youth, would the same motivations be present for adults to commit criminal offenses? Haddad & Moghadam (2011) showed that economic factors such as legal and illegal payoff opportunity were significant for crime against properties and threat. This study was able to determine that educational factors were considered to be a main determinant. With the lack of education a person is not able to attain the level of skill or knowledge needed to progress as a person might would anticipate. As unemployment increases during a country's economic growth, crime will become more prevalent based on the growing capitalism and population.

Snook, Dhami, & Kavanagh (2011) supports Haddad et al. (2011) when they observed the choices criminals make when they do commit a crime. Usually a fast and frugal decision is made using some form of heuristic strategy that is often noncompensatory. Whatever crime that may be committed is mostly committed out of what the criminal believes is necessary given the current situation and their perception of the situation in regards to how they plan on to react.

Another study found that when interviewing burglars there was an excitement factor to committing the crime (Nee & Meenaghan, 2006). That merely committing the act as well as the rewards for being successful was a contributing factor to executing the criminal task.

Specific criminal activity.

Through research generic reasons were cited why a person would commit some form of criminal act. Based off the literature reviewed majority of support goes towards early conditioning and motivation based on the perceived necessities of the criminal. So how does a criminal, such as a burglar, determine when and where they will target a residence?

Nee & Meenaghan (2006) showed that burglars operated using an almost habitual decision-making process that all experienced individuals use to navigate quickly and effectively around their world. Burglars discussed the need for money was the primary motivating factor in the initial decision to burgle. Also discussed in the same study, the shear excitement of committing the crime was a powerful motivator as well.

Snook et al. (2011) observed that burglars utilized characteristics of the victim of the residence they intended to target. As reported by burglars in Snook et al. (2011) the presence of the victim at home was the most important factor to consider as well the presence of a security system.

Pitcher & Johnson (2011) demonstrated that a burglar’s action is more of a forging and heuristic process when selecting targets. Usually the burglar is looking for a combination of factors that would lead them to believe the target is possible. Such considerations are the specific period of time for the operation, selection of the target, whether or not the target has been successful in the past and what the terrain is in regards to concealment, ingress and egress.

Nee et al. (2006) learned that having prior knowledge of the target; its occupants and the potential profit from that target did not weigh too much into an offender's decision. The initial draw to the target is the general upkeep and décor, visible expensive items, and the type of car present. Nee et al. further determined that little preparation was used and prior knowledge of contents of a residence and the activities of the occupants were observed by the burglar.

Rengart (2004) determined there were more characteristics to the burglars target selection such as minimizing the commuting distance; whatever the closest opportunity might be. The burglar is taking other factors not directly associated with the target to weigh even further to their target selection. Such factors would be the amount of distance from their home base to the target for travel time and security considerations; will they need to provide transportation, will they be seen, what items can I take? Elffers (2004) supports this thought process while interviewing prior offenders and learned natural time constraints, high cost of reaching remote targets, and the limited information that was available on remote locations. Canter and Hammond (2006) showed that distance is also relative to the offender and not so much defined by the actual characteristic of distance. Ultimately it was up to the offender to determine how far they were willing to travel to commit an offense.

Fritzon (2001) showed that criminals utilize distance as a buffer or safety zone when selecting targets. Burglars are more apt to select targets at a "medium" distance so as not to operate near their homes or along active patrol routes. This would decrease the likelihood of recognition or arrest. In the case of Bernasco & Nieuwbeerta (2005) it was determined that neighborhoods being selected for burglary is positively influenced by the apparent lack of guardianship. In other words, their decision to commit the offense was dependent on the access of the neighborhood and the patrols of security personnel and police.

Another factor that weighed into an offender's travel pattern was also the terrain itself, more specifically roadways and manmade and natural barriers. Greenberg and Rohe (1984) showed that low-crime neighborhoods were limited to outside access to heavily traveled roadways. Peeter (2007) demonstrated further that different types of barriers have different impacts. For example a major road way might be considered a barrier in the event it needs to be crossed to meet the burglar's egress. However if it became part of the offender's egress route it is no longer a barrier. Clare, Fernandez, & Morgan (2009) showed there is a significant effect of connectors and barriers influencing target selection in that based on the amount of impermeable barriers in relation to the proximity to the offender's origin has an impact in target selection.

Burglars were shown to more likely target homes in close proximity to homes targeted in the past (Townsley, Homel, & Chaseling, 2003; Sagovsky & Johnson, 2007). This was seen with homes that had little architectural differences from previously targeted homes in the past. This falls in line with repeat burglary victimization. Bernasco and Nieuwbeerta (2005) demonstrated exact repeat and near repeat burglary patterns are often explained in terms of event dependence, that burglars preferred target familiarity.   However the risk for repeat victimization was not at a constant high risk, but was related temporally and that the likelihood of repeat victimization was highest within a short period of time, as the distance in time increased the risk reduced. Sagovsky and Johnson. (2007) showed that repeat victimization risk reduces swiftly during the first four weeks that follow the initial crime.

Patterns of the crime.

Utilizing information attained from research and historical data, researchers and law enforcement were able to develop means to target the burglar utilizing the characteristics of the crime, location, time of day, and day of week (Ratcliffe 2004; Ratcliffe and Mcullagh, 1998). By observing these specific factors the criminal offense starts to collect together into what has been termed "clusters" which are based on a combination of spatial and temporal characteristics (Ratcliffe, 2004). Johnson, Bernasco, Bowers, Elffers, Ratcliffe, Rengent, and Townsley (2007) utilized data from five different countries and found the same effect of clustering based on the same characteristics and showed more burglaries occurred close to each other in space and time.

Clusters or "hot spots" are generated in an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization (Eck, Chariney, Cameron, Leitner, and Wilson, 2005). Crime hot spot maps have effectively guided police action guided by crime theories, but not all hot spots are the same. Hot spots have to be observed in a spatial context in comparison to the areas they represent. There are three theories that are observed when considering spatial reference to crime. Place theory focus on why crime events occur at specific locations or the lowest observable level such as residence or businesses. Street theory is when events that span over small stretched areas such as blocks and streets. Research has shown that there is an interaction of targets and offenders along thoroughfares in relation to going to and from work, recreation, shopping, and school. Finally neighborhood theory focuses on largest scale analyzing square blocks, communities, and census tracts.

To support the spatial method theories is additional theories such as repeat victimization and crime opportunity. Repeat victimization is the likelihood of a residence to be targeted again within a given time period. Crime opportunity is the likelihood of an offender committing a crime along a stretch of area constantly traveled, but with no planning or targeting. Both coupled with the Place, Street, and Neighborhood theories analysts can manipulate maps depicting crime to produce or disperse hot spots (Eck, Chariney, Cameron, Leitner, and Wilson, 2005).

Based on factors associated with offender traveling distance, incorporating a distance decay function to determine if there was a link between point of origin and the location of the offence as well as to what extent the spatial patterns of property crime can be explained by income inequality and proximity. Rattner & Portnov (2005) observed the incidence of property crimes decline as a function of distance separating delinquency areas and places of criminal residence. It was determined that property crime in major cities of Israel are committed by offenders who live outside these cities with criminal presence dropping steadily as the distances between places of criminal residences and the central city increase.

Other methods to determining criminal behavior.

Spatial and temporal analysis has been the primary means of determining the epicenter of criminal activity solely based on the environment and the impact of the behavior upon the environment. Jing and Tao (2011) developed a decision tree classification utilizing an algorithm to reduce the extraneous behavior available not associated with a particular type of crime. The researchers were able to generate significant results based on their algorithm, but also identified that the attributes associated with criminal behavior must be finely tuned to provide the best results.

Pillai and Kumar (2007) utilized clustering algorithms to determine how many agents were present given data representing "hot spots" of criminal activity. The study broke down the environment to represent three levels of decisions based on the activity of the environment;

presence of law enforcement, distances associated with the target, and the opportunity to commit the offense. Based on when a signal was registered given which environmental factors associated with the signal the researcher could distinguish how many agents were active within a given time period. However this method can only be utilized to predict the likely area that will be targeted next and also relies on a massive amount of data to train the model correctly. Moreover the tool can be utilized in “real world situations”.

Ferrari, Baumgartner, Palermo, Bruzzone, and Strano (2008) conducted a study to determine if the environmental variables associated with a crime in conjunction with psychological characteristics of the criminal could be used to predict the when a criminal action would take place. The researchers utilize neural networks and Bayesian network models of the criminal behavior developed from a database which categorized the psychological behaviors of criminals based on the offense committed. The study was able to pull information from cleared case databases and reduce or eliminate suspects be comparing the factors within the cleared cases with the ongoing investigations profiling the activities of potential suspects. Additionally they determined by conducting a sensitivity analysis the researcher was able to determine the significant relationships among the observed variables.

Linear Regression

Linear Regression is the study of the dependence one or more variables has on a particular variable with the goal to summarize an observation as simply, usefully, and elegantly as possible (Weisburg, 2005). More specifically, linear regression attempts to model the relationship between a designated set of independent variables by fitting a linear equation to the observed data (“Linear Regression”, 2011). Being one of the oldest and most widely used predictive modeling means, linear regression minimizes the sum of the squared errors to fit a straight line to a set of data points. The intent is to determine, based off a given data set, a specific relationship of the predictors (independent variables) and how it affects the outcome (dependent variable) of the developed model. If there was a best fit between the function and data, the actual value of the target value for each record in the data would equal the predicted value.

This method is initially conducted by determining whether or not there is a relationship or significance between the variables of interest by developing a scatter plot. Be doing this we are able to pictorially depict and determine if there is a general linear relationship between the independent variable (IV) and the dependent variable (DV). If it appears there is no association between variables, fitting a linear regression model to the data will not provide a useful model. The regression would not be characteristic of the behavior of the observed behavior. Also the scatter plot will denote some form of correlation coefficient or what type of and strength of the relationship between the variables. For example if the correlation coefficient was a value of -1 or negative value there is an inversely related relationship depicted by a downward slope, where a value of 1 or positive value suggests a direct relationship between the variables depicted by an upward slope (“Linear Regression”, 2011).

The most common form of fitting a regression line is the Least-Squares Regression or Ordinary Least Squares (OLS). This method calculates the best fitting line by minimizing the sum of the squares of the vertical deviations from each data point to the line. Ultimately the OLS will determine, regardless of the data set, the best fit line that depicts the linear relationship between the IVs and the DV. The disadvantage of the OLS is the limitation in the shape the linear model can assume over long ranges, due to poor extrapolation, and sensitivity to outliers

("Linear Regression", 2011). Essentially the best fit line would not be characteristic of the actual behavior of the observed factors.

Basic equation.

The equation normally associated with Linear Regression is as follows:

$$Y = a + bX \tag{1}$$

The X is the explanatory variable, independent variable (IV), or predictor. It is the contributing factors researchers identify and proceed to manipulate within a model. Y is the dependent variable or what the researcher means to record based on the interaction of the IVs within the model. The b refers to the slope of the line or the general relationship the independent variables has on the dependent variable as the independent variable proceeds from 0. The a refers to the intercept, value of Y when $X = 0$ or the natural state of the dependent variable prior to the interaction with the independent variable ("Linear Regression", 2011).

Different types of functions.

Based on the number of predictors the function will depict a line of best fit. In the case of a single predictor the function will describe a straight line fit. With the introduction of another predictor (two) the function will display more of a plane incorporating the effects of more than one IV and their interaction with each other. With multiple predictors the function will then represent a hyperplane to the n-dimension depicting the relationship between all the variables and it affects upon the dependent variable (Sherrod, 2011).

Outliers and influential observations.

If there is a difference between the actual value of the dependent variable and its predicted value for a particular data set, then there is an error of the estimate or commonly known as the deviation or residual. This is easily explained as the difference between the data above and below the line of best fit. This type of effect can be attributed to a data set as it is presented which may include outliers and influential observation. Outliers refer to erroneous data that if unchecked or unaccounted for will result in a poorly fit regression line. Influential observations are data that lies far from the other data along the horizontal direction; this too will have a profound effect upon the regression line by affecting the slope. Residuals is the deviation from the fitted line to the observed values, so depending on the distribution of the data points may depict a linear relationship but so scattered in relation to the best fit line may yield a low significance (“Linear Regression”, 2011).

Residuals are not only representative of poor data but can have an effect upon validity of the study. However, by investigating the residuals the researcher may reveal a non-linear relationship within the study, which can be attributed to Lurking Variables. A lurking variable is a third variable not directly measured that has a significant effect upon two observed variables within the study in turn affecting the study if not addressed and accounted for. A useful tool to identifying lurking variables is the use of a Time Series Plot. The time series plot will utilize the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values in turn identifying the lurking variables in relation to other data points plotted (“Linear Regression”, 2011).

Strengths and weaknesses of linear regression.

Linear regression is one of the most widely used methods in modeling and predictive analysis and it is well understood. Manipulating a linear regression model is usually much faster and is simple and requires minimum memory for computer systems to implement. Also by examining the magnitude and sign of the regression coefficients it can be infered how predictor variables affect the target outcome (Sherrod, 2011).

Along with strengths there are weaknesses associated with linear regression. Multicolinearity which refers to the effect of more than one predictor if used in the model and potentially strong relationships between the predictors which can result in large errors in calculating the parameters of the model can be problematic. Lastly, if the data presented represents more of a skewed function it will not represent a typical function of a regression model (Mason and Baddour, 2007).

Assumptions when utilizing regression models.

When determining to utilize regression models the researcher must make some assumptions prior to initializing their study. The first assumption is that the independent variables must be in interval, ratio, or dichotomous format in nature as well as the dependent variable must be continuous, unbounded, and measured on an interval or ratio scale. More specifically the data needs to be accurate and free of errors from the time of data collection (Menard, 2002).

Another assumption is that relevant predictors of the dependent variable are included in the analysis, no irrelevant predictors are included, and the relative form of the function must be linear. Studies utilizing linear regression must narrow down the IVs precisely and focus on the IVs researchers wish to manipulate. Independent variables that have no bearing on the study may produce outliers, skew the best fit line, or develop lurking variables that will not be able to explain if not anticipated for prior to the study (Menard, 2002).

Regardless of the data utilized, the IVs manipulated, and the DVs recorded, the expectation of the study will ideally yield an expected value of error of 0. This states the IVs utilized within the study must accurately depict the observed data with no errors. As error is introduced, the resulting effect is an increase in deviations from the line of best fit, and the less significance the results will represent the linear relationship of the predictors and target variables (Menard, 2002).

The IVs must have homoscedasticity or the variance of the error term is the same or constant for all IVs. In other words the variance between the variables is relatively the same and will have no impact upon the relationship they share towards the DV. If the variance is significantly different between the IVs, the results will be indicative of one IV having influence over other IVs utilized in the model (Menard, 2002).

Errors are assumed to be normally distributed for each set of values of the IVs, resulting in a normality of errors. In other words the normal distribution of real valued random variables that tend to cluster around a single mean value must be representative of the population in proportion to the sample size (Menard, 2002). This is another assumption of applying linear regression when determining a relationship between predictors.

There must be no correlation among the error terms when different IVs are entered within the function. This states the function will yield different results when different IVs are utilized. If the results do yield the same values regardless of the IVs utilized, then the IVs have no effect upon the function whatsoever (Menard, 2002).

The function must also be absent of perfect multicolinearity when observing multiple regression functions. None of the IVs plotted must yield a perfect linear combination, this may implicate there is no interaction between the IVs regardless of the combination utilized within the study to yield an effect (Menard, 2002).

Logistic Regression

Logistic Regression is similar to linear regression but is best suited to conduct analyses with a continuous target variable in which the prediction function yields a value of 1 for one category and 0 for any other category. An example would be a study where one must determine to pay for an item based on the variables or not to pay for the item. There are two distinct outcomes that are expected to be yielded from the study, one which supports the target variable or buying the item and the other which does not. Ultimately logistic regression is interested in describing or testing a hypothesis based on the categorical outcome of independent variables.

The intent of the logistic regression given the IVs; the DV will yield a value that represents a relationship of the IVs within a given situation. Also within logistic regression the DV can be represented as "1" or "0" to describe a specific category in what is being studied. Unlike linear regression where the *Y* value can fall anywhere from "-1" to "1", and require detailed analysis, logistic regression will simply state based on the provide IVs what the effect will be. This will also not yield a linear function, given there will only be two answers to describe the interaction between the IVs.

Logistic regression function and model.

The logistic regression function is designed to denote a probability that will lie between the values of 0 and 1 regardless of the magnitude of the *z* as it can range from negative infinity (-∞) to positive infinity (∞). The logistic regression function:

$$f(x) = \frac{1}{1 + e^{-z}} \tag{2}$$

The logistic regression model computes the probability of the selected response as a function of the values of the predictor variables. The formula is as follows:

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots . \beta_k X_k \tag{3}$$

z is the index that combines the observed independent variables. X_k represents the variables that are of interest to the researcher, the independent variables. β_k represents the constant utilized within the function which is unknown, as well as α, which represents the Y intercept. The Y value corresponds to the outcome of interest. The conditional probability of the dependent variable can be expressed by a logistic model as shown:

$$P(Y = X_1, X_2 \dots . X_k) = \frac{1}{1+e^{-(\alpha+\Sigma\beta_k X_k)}} \tag{4}$$

Utilizing this model, a researcher can ascertain how well it predicts probabilities, or in the accuracy of prediction of a particular group. These classifications are broken down into three separate types of models; prediction, classification, and selection to determine the best representation of the study and data (Kleinbaum, Klein, and Pryor, 2010).

Prediction, classification, and selection models of Logistic Regression.

Prediction models are not constrained to the number or proportion of data in order to predict to have or not have a specified behavior or characteristic. The study is merely trying to determine if the model will make a decision, based on the model's IVs, to choose one option over the other based on the provided data. This method is appropriate when the data being tested and the IVs utilized are utilized in the same respects with no manipulation.

Classification models imply the data are heterogeneous and requires the model to classify all the provided data points into specific categories as specified by the researcher. Because of the established criterion, the data has to be sorted in a manner that best meets the provided categories. The system does not have the discretion to establish a new category if the criterion for the established categories are not met and must select the category of best fit.

Selection models are concerned with the acceptance or rejection for inclusion in a specified group which satisfies a criterion for success in a particular category minimum required, maximum allowable, or specified number of groups allowed by the researcher (Menard, 2002).

Evaluating a Logistic Regression model

Once the model has been constructed, it is necessary to evaluate if the model can be utilized for its intended use. In order to evaluate the model the following four areas must be addressed to validate the model; overall model evaluation, statistical test of the individual predictors, goodness-of-fit statistic, and validation of the predicted probabilities.

An overall model evaluation must first be conducted to determine if the model provides a better fit to an observed data set than if the data was applied to an intercept-only model. The intercept-only model, which can be referred to as a null model, serves as a baseline for comparison due to the lack of predictors. Any improvement observed in comparison to the baseline will be examined utilizing inferential statistical test such as a likelihood ratio, score, and Wald tests. Not in all cases these three tests will yield the same results, in such a case rely solely on the likelihood ratio and score tests (Peng, Lee, and Ingersoll, 2002).

A statistical test of individual predictors must be conducted to determine the significance of individual regression coefficients. This will notify the researcher whether or not the predictors they have chosen are significant enough to produce an effect. The test of the intercept should be conducted to determine if a the intercept chosen can be utilized within the model. These evaluations can be conducted by utilizing a Wald chi-square statistic (Peng, Lee, and Ingersoll, 2002).

A goodness-of-fit statistic can be utilized to assess the fit of a logistic model against actual outcomes of the model. This is accomplished by utilizing either one inferential statistic test and/or two descriptive measures. The inferential option is conducted utilizing the Hosmer-Lemeshow test which is a Pearson chi-square statistic which is derived from a 2 x *g* table of observed and estimated expected frequencies, where *g* is the number of groups formed from the estimated probabilities. Criteria that should be met when utilizing the Hosmer-Lemeshow test is an equal number of observations, the number of groups should exceed 5, and expected frequencies should be at least 5 (Peng, Lee, and Ingersoll, 2002).

The two descriptive measures of goodness-of-fit are R^2indices; Cox and Snell (1989) and Nagelkerke (1991). The indices describe a proportion of the variation in the dependent variable that can be explained by the predictors utilized in the model. These indices do not render the meaning of the variance explained, corresponds to predictive efficiency, or can be tested in an inferential framework. The researcher is forced to treat the two indices as supplemental information to the overall evaluation of the model, tests of individual regression coefficients, and the goodness-of-fit test statistic (Peng, Lee, and Ingersoll, 2002).

Finally the predicted probabilities must be validated to ensure the model meets the intended outcome. The resultant predicted probabilities can be revalidated with the actual outcome to determine if high probabilities are indeed associated with events and low probabilities with non events. The degree to which predicted probabilities agree with actual outcomes is expressed as either a measure of association (Kendall's Tau-*a*, Goodman-Kruskal's Gamma, Somer's *D* statistic, and the *c* statistic) or a classification table, which is utilized when classification is a stated goal of the analysis (Peng, Lee, and Ingersoll, 2002).

Studies Utilizing Linear and Logistic Regression

Weisz, Tolman, and Saunders (2000) utilized linear regression to study domestic abuse in relation to the amount of physical or verbal contact of the victims to determine the greatest severity of abuse. The study observed demographic data to determine if there was a correlation between victim and perpetrator demographics and a relationship between the two. Additionally the location of the incident being a residence, on the street, or miscellaneous location observed from police reports. Regardless of location or severity, the study was able to determine the majority of domestic abuse appeared to be clustered between the hours of 9:00 p.m. and 11:00 p.m. during any day of the week.

Law, Shapka, and Olson (2010) observed the online aggressiveness of young adults in relation to parent interaction and availability of a personal computer. The study determined not controlling or monitoring online activities was negatively associated with online aggression and the young adult was more likely to engage in aggressive behavior. The study also yielded having access to a computer, i.e. laptop in the young adults bedroom, yielded a higher probability of engaging in online aggressive behavior. Only adult interaction, i.e. engaging the young adult about online activity, and lack of access to a personal computer showed a higher probability from deterring online aggressive behavior.

One study utilized linear regression to identify battery usage attacks by observing characteristics associated with a computer's battery draining excessively when the system is performing effectively due to battery usage attacks. Nash, Martin, Ha, and Hsiao (2005) identified the combination of CPU load, disk read and write accesses, and network transmits and receive function contributed to a battery usage attack. Additionally the study observed types of battery usage components such as display consumption, CPU fan, and other devices not monitored directly by quantifying them utilizing correlation coefficients. The system was able to determine based off the criteria when the CPU was experiencing a battery usage attack.

Curtin (2011) utilized linear referencing to identify characteristics along a road network to determine the probability of an Improvised Explosive Device (IED) being emplaced along a road network. The study observed characteristics such as intersections, culverts, and urban/ rural events as predictors in relation to their distance and interaction with a particular road to determine the likelihood of an IED would be emplaced along that road. Curtin was able to predict the likelihood of an IED being emplace along a particular road based on the combination based on the aforementioned characteristics to assist in identifying a potential Targeted Area of Interest (TAI) for Coalition Forces (CF) as well as the predictors that influenced the emplacement of an IED such as urban/ rural activity or CF presence.

Rationale for Study

Criminal behavior and criminal target selection in terms of burglary is believed to be predictable by certain characteristics. By identifying these characteristics and analyzing their influences in the decision making process of the criminal we can develop a means to identify potential targets or general locations the criminal will attempt to target in the future. In turn this will allow law enforcement professionals to focus resources based on the analysis, exploiting key information necessary to reduce or eliminate the criminal threat. Additionally there are abundant studies that utilize spatial and temporal factors, but very little studies represent a combination of factors as discussed earlier.

This research study will seek to determine if targeting by a burglar is dependent on specific conditions that satisfy the burglar's targeting decision making process by using logistic regression. Prior research supports time and location can be utilized to depict the high activity, as well as repeat victimization, as well as relation to connectors and barriers. Prior research has not supported a combination of all the aforementioned variables in the determination of criminal behavior. Burglars target a residence based on a specific time and location based on repeat victimization and distance to prior targets, and the influence of connectors and barriers in their decision making process. Regardless of the combination of these factors, burglars simply utilized forging and holistic behavior to determine when to target a residence.

Method

Problem Formulation

This study was designed to analyze and predict burglar behavior utilizing logistic regression to determine the pattern with regards to a set of individual variables associated with burglary. Due to the rise and fall of the economy as well as the availability of jobs in a tight economy, burglary and theft have become a means to satisfy financial shortcomings to meet personal needs and wants (Haddad & Moghadam, 2011).

Utilizing variables from previous research, a logistic regression model was developed to predict the pattern of a burglar based on the target selection factors within a specified area. The studies reviewed have utilized similar methods to develop cluster or pattern analysis to depict commonalities in targeted residences. Commonalities such as time of day, day of the week, and location in relation to past burglary events have been utilized in pattern analysis (Ratcliffe & McCullagh, 1998). Similar research utilized different variables to determine if there were other factors associated with the targeting of particular residences or neighborhood. Factors such as repeat victimization (Short, D’Orsogna, Brantingham, & Tita, 2009), likelihood of being targeted again, and the effects of barriers and connectors (Peeters, 2007), natural or manmade obstacles or pathways all have had an effect on targeted locations. This study analyzed these factors as predictors within a logistic regression model, to determine if a burglar's pattern can be more accurately predicted, by using this model.

Data Collection

The data consisted of consolidated information from filed police reports from the year 2010 by the City of Daytona Beach, Florida and the Daytona Beach Police Department. The current crime rate of 7.70 percent resulting in one in thirteen people of having the likelihood of being victimized via burglary, theft, or motor vehicle theft within the city concerning property crime provides enough data to support the study. Daytona Beach's property crime rate is 42.34 percent higher than the national average, which makes it a proper test bed for this study.

The data was compiled from reports generated from police officers when responding to a reported incident. The Daytona Beach Police officer takes note of the information pertinent to the crime, for burglary; time of the police response, time of the offense, location, items missing, persons interviewed to include neighbors or associated personalities, method of entry, method of egress, and additional forensics conducted by the police officer. Towards the end of the shift the officer translates his notes into a specified form for the offense and provided to the departments clerks for documentation and populated into the criminal database. The analyst for the police department reviews the actual inputted data versus the official form to clarify or correct deficiencies when the information was entered into the database.

The information was provided in a Microsoft Excel spreadsheet format and transferred from the Daytona Beach City Records division. The spreadsheet consisted of specific information based on prior research discussed in the introduction of this thesis. The information provided by the city is all reported and investigated burglaries within the Daytona Beach Police

Department jurisdiction. The data was comprised of the time and date of the offense, approximate start and finish time of the offense, and the street address of the offense.

Variables.

Independent variables.

The independent variables for this study are based on prior studies in crime mapping and analysis. The first variable utilized is time of day, more specifically, the specific time the offense was believed or reported to have occurred. As Ratcliffe (2004) demonstrated that time has as much an effect upon target selection as does the location and specific periods of the day were targeted for various reasons.

The second variable, the specific day of the week, was recorded and assigned a category value to depict which day of the week is represented within the data; for example, Sunday – 1, Monday – 2, Tuesday – 3, etc. Studies such as Ratcliffe (2004) demonstrated that the particular day of the week was as important as the time of day. The criminal is most likely to target a residence when the criminal believes the resident and neighbors are not likely to be home due to victim and neighbor work or daily life habits.

The third variable, repeat victimization is the occurrence of the offense at the same residence following the initial offense over the calendar year. Bernasco and Nieuwbeerta (2005) demonstrated residences were more likely to be victimized again based on the criminal's familiarity of the layout of the target and items desired by the criminal within the target. The data provided will be coded to indicate which residence was a repeat victim; 0 – Not a repeat victim, 1 – repeat victim.

The fourth variable, connectors, refers to the amount of access streets, pathways or bridges relative to the targeted residence that allows access to a multi-lane street. This was done by observing the actual residence and identifying the connectors as if a person exited the front of the house and attempted to move forward, left, right, or through the rear of the residence. Barriers, the last variable, will be counted as physical structures, underdeveloped pathways, and street dead ends that will disrupt or block an individual's egress from a targeted residence from reaching a major street system or adjacent neighborhood in the same relation associated with connectors (Peeters, 2007). Table 1 illustrates the data format of the independent variables.

Table 1
Independent Variable Description

Independent Variables	Data Type	Description
Time	Categorical	Time categorized into three groups (0800-1600 – 1, 1601-2359 – 2, 0001-0759 – 3)
Day of Week	Categorical	Value assigned to represent the day of the week (SUN – 1, MON – 2, TUES – 3, WED – 4, THURS – 5, FRI – 6, SAT – 7)
Repeat Victimization	Categorical	0 – Not repeat victim, 1 – Repeat victim
Connectors	Discrete	Number of access routes (streets, pathways, bridges) connect to multi-lane streets or adjacent neighborhoods
Barriers	Discrete	Number of physical structures, underdeveloped pathways, and street dead ends that disrupt or block a major street system or adjacent neighborhood

Dependent variables.

The dependent variables utilized for this study reflect the values that were produced utilizing the binary logistic regression function and dictated by the distance from the crime epicenter. Location was the actual residence, in which the offense took place, and was categorized by determining the events that fall within incremental radius from an identified concentration of events ranging from 1 to 9km. Hammond and Young (2011) demonstrated that crime activity will reduce as it increases in distance from the identified epicenter of activity. The function is designed to depict whether a signal is registered based on the combination of independent variables within the function. In the event the variables do depict a relationship with the act of committing a burglary, within 1 km increments of the crime epicenter (Hammond and Youngs, 2011). Crime within the different 1 km incremental increases was compared to one another as they progress from the epicenter. The purpose is to identify the predictors' relation in comparison to the radii they fall into and the strength of that relationship increases from the epicenter. The function will only be limited to the two responses based on its purpose and design in registering a signal. Table 2 illustrates the dependent variables.

Table 2
Dependent Variable Description

Dependent Variable Description	Description
Events < *n* km radius	Depiction of associated criminal events
Events > aggregate km radii	Depiction of non associated criminal events

The variable *n* represents the distance of the km radii. Example: < 1 km in comparison to > 1 km, < 2 km in comparison to > 2 km, < 3 km in comparison to > 3 km.

Model of study.

The following equation represents the variables utilized to determine the probability based on the interaction of the observed predictors of time of day, day of week, repeat victimization, connectors, and barriers.

$$z = logit\ (Burglary\ Activity) = \alpha + \beta_1(\text{Time of Day}) + \beta_2(\text{Day of Week}) + \beta_3(\text{Repeat Victimization}) + \beta_4(\text{Connectors}) + \beta_5(\text{Barriers}) \tag{5}$$

Procedure

The data was first extrapolated based on the police reports provided to determine the variables for this study's scope; repeat victimization, day of the week, connectors and barriers. The date was observed and cross referenced with a calendar to determine which day of the week the offense took place. The street addresses were mapped and the concentrated epicenter was determined by developing a grid reference system, annotating the amount of events per grid, and mathematically determined based on the concentration of events within each grid. Incremental diameter circles of 1km were applied from the epicenter to categorize the distance from the established epicenter. This was accomplished by plotting every address in Google Earth and Maps effectively providing the pictorial depiction of the data. A graph was placed over the concentrations to determine how many events take place within each grid of the system. The data was inputted into Mathematica to determine the concentrated epicenter of the events utilizing the following formulas:

$$(i) = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} i\ v(i,0)}{\sum_{j=1}^{n}\sum_{i=1}^{n} v\ (i,0)} \tag{6}$$

$$(j) = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} j\ v(i,0)}{\sum_{j=1}^{n}\sum_{i=1}^{n} v\ (i,0)} \tag{7}$$

Once the addresses were mapped in Google Earth and Maps the researcher can determine the amount of barriers and connectors associated with each targeted residence. This was accomplished by observing the amount of streets that would allow an individual to move to a major street or adjacent neighborhood system unopposed. A major street system was defined as a multi-lane street with or without a physical medium that can be accessed by turning immediately with the flow of traffic as governed by the local traffic laws. Adjacent neighborhoods were defined by neighborhoods that can be accessed via foot path or connecting residence street systems. Finally it was determined which residences were repeated targets within the given year. This was accomplished from observing the data and annotating which residence which were targeted more than once over the course of 2010.

Once the data were collected, analyzed, and annotated the information was inputted into SPSS in order to run inferential statistics to determine if any of the variables were significant ($p < 0.05$) prior to conducting a binary logistic regression. A binary logistic regression was conducted utilizing a majority portion ($N = 600$) of the information to develop a regression model to determine the probability of observed variables and their relation to a burglary event. The variables were also observed in conjunction with one another as well as specifically paired to determine if there is any specific interaction in determining predictability. Once the probabilities for each variable were collected, SPSS was utilized in conjunction with the remaining data set and the probabilities established by the initial statistical run to determine if the variables can predict if another event will take place within a given area.

Validation and Verification

Statistical Validation of Logistic Model.

The logistic regression for the model was evaluated as per the referenced guidelines by Peng et al. (2002) when evaluating a logistic regression model. The model was first compared to an intercept-only model, which will be considered a null model, to determine if the observations would be predicted to belong in the largest outcome category. Then inferential statistics (likelihood ratio, score, and Wald statistics) were utilized to determine to what degree the model performed in relation to the intercept-only model.

A Wald chi-square statistic was utilized to determine the significance of the individual regression coefficients and whether or not they are having some effect upon the regression model. A Wald chi-square was utilized for the intercept to determine the significance of the intercept affect upon the model and whether or not the model should utilize the intercept.

The goodness-of-fit statistic to determine the fit of the logistic model was done utilizing the Hosmer-Lemeshow test which is an inferential goodness-of-fit test. R^2 indices were utilized to supplement the Hosmer-Lemeshow test to determine the proportion of the variation in the dependent variable that can be explained by predictors in the model.

The resultant probabilities from the model were then revalidated with the actual outcome to determine if the high probabilities are associated with the events in comparison to the incremental one km radii. This was determined by utilizing a measure of association, more specifically a Somer's *D* which is the extension of a Goodman-Kruskal's Gamma, which utilizes the dependent variable and one of the independent variables and represent the degree of association between the outcome (designated dependent variable) and the estimated probability (independent variable).

Google Earth and Maps.

Google Earth and Maps were utilized to extrapolate additional information that cannot be provided by the City of Daytona Beach. Google Earth and Maps is an interactive online geographical tool that was originally designed to depict the Earth's surface utilizing satellite imagery, satellite global positioning, and established infrastructure and architecture to depict the current state of the Earth. The program allows an individual to enter a place, as specific as dictated by the user, and move directly to that location with a minimal degree of error. The program is utilized to find specific street addresses, business, geographic coordinates (military or nautical), landmarks, cities, countries, and all major land features stored in the Google server system. The program's map view does not become fixed once a desired location is input into the system. The user has the ability to adjust the magnification as close as a first person view, if the Google has photographed the particular area. If no first person view is available for a given area the program will default its closets detailed image to above street level which can clearly depict cars, streets, vegetation, and details associated with residences and established infrastructure.

SPSS program.

SPSS was utilized based on the widespread use and acceptance of the programs abilities to conduct multiple types of statistical analysis in support of research studies. SPSS statistics takes data from almost any type of file and use them to generate tabulated reports, charts and plots of distributions and trends, descriptive statistics, and complex statistical analyses.

Utilizing the binary logistic regression function in support of this study provided face and content validity to the analysis. The binary logistic regression function utilizes whether the established variables correlate with observed action, in this case burglary. During the prediction phase of this study, the association of a group of burglaries within a 1 km radius was registered a value of 0 and those associated with burglaries within the next incremental km circle will register a value of 1. There was no other alternative to the association of a group of burglaries or no additional association for the program to select based on the provided dependent variable.

The study's internal and external validity was dependent upon the information provided by the City of Daytona Beach, the researcher, and selected programs. The information provided by the City of Daytona Beach is as accurate as the reports generated by the Police Officers who investigated each report, the clerks who entered the information into the police department database and cross reference of the analyst prior to delivery to the researcher.

The researcher utilized established parameters of selection of connecters and barriers and as established in prior research in conjunction with a mapping program to extrapolate information pertinent to the study; population of burglary events and their geographical position in relation to the concentrated group epicenter and the presence of connectors and barriers. All this information is dependent on the accuracy of the mapping program and the current information within the system itself.

The variables within this study, through separate research, have been utilized to develop predictive models in prior studies. The information can be easily attained through request from a local law enforcement agency and cross referenced with public records. The information provided by the mapping program can be verified by traveling to the actual destination dictated by a person based on the data set provided for the study. The statistics can be verified by hand based on established mathematical functions and equations as discussed in the literature review. The study can easily be reproduced utilizing the same instruments, procedures, and data set within the same area or any other location that maintains records concerning the topic or similar topics, access to a computer, and the use of a mapping and statistics program.

Experiment Objective

This study explored if the combination of variables discussed in the Introduction and Methods sections of this thesis can accurately predict the action of burglary within a given area. Studies have shown time and location has been a big determinant when targeting a burglary act, but also other environmental factors have depicted accuracy as well. This study aimed to incorporate all those aspects not previously tested utilizing a logistic regression model. Specifically different comparisons of radii were utilized to project a pattern of crime as it projects from the epicenter of activity. If the model is effective in predicting the committal of a burglary action based on probability, the model can be utilized and expanded to incorporate other forms of criminal behavior. Such behavior as vehicle theft or vehicle burglary utilizing the similar theory established in this study can be utilized to predict the probability of an action being committed. The scope can be expanded beyond law enforcement and incorporated into military application as well. Utilizing similar data as supported by research, predicting the likelihood of an Improvised Explosive Device (IED) utilized on Coalition Forces (CF) during combat operations can be exploited. Such information will direct and target Route Clearance assets, Intelligence, Reconnaissance, and Surveillance (ISR) assets, and combat patrolling to disrupt, capture, or deter IED emplacement.

Results

A logistic regression analysis was performed to determine burglar behavior as it relates to distance from an established epicenter of activity and five associated predictors: repeat victimization, time of day, day of week, presence of connectors and presence of barriers. Analysis was performed using SPSS binary logistic regression. Six hundred documented cases of burglary activity were partitioned and compared based on their distance from the identified epicenter. The model generated for this study is as follows:

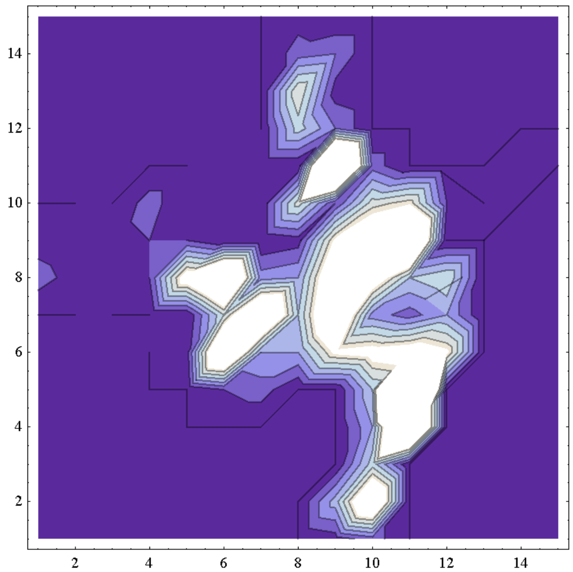
$$z = logit\ (Burglary\ Activity) = \alpha + \beta_1(\text{Time of Day}) + \beta_2(\text{Day of Week}) + \beta_3(\text{Repeat Victimization}) + \beta_4(\text{Connectors}) + \beta_5(\text{Barriers}) \quad (8)$$

All the events were categorized based on location within the Daytona Beach area by developing a 1 x 1 grid system to quantify the amount of activity within a one kilometer area to determine the overall concentration and epicenter of the events. The results are illustrated in Figure 1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1									1	3					
2									7	21					
3										9					
4										6	30				
5						2	3			8	57				
6					1	23	6	6	14	15	16	14			
7						10	23	8	21	5	3	6			
8	3	1		2	16	25	5	6	19	17	8	10			
9	1		1	2	2			1	14	19	27				
10				3		1		11	2	11	13	2	2		
11									34	3				1	
12								9	4						
13								12	1						
14								3	4						
15								1							

Figure 1. Number of burglary events in relation to the grid system utilized to categorize events to determine the concentration and epicenter of events. Each grid square represents a one kilometer area (1 km x 1 km).

The results of the categorizing were utilized with Mathematica to develop a density plot as well as determine the mathematical epicenter of the events. This was done by calculating the epicenter needed to use the merge values of the coordinates with the weight v(i,0). The epicenter was determined to be at 9.10 units along the x-axis and 7.30 units along the y-axis of the grid system. The graphical results are shown below in Figure 2 and 3.



$$(i) = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} i\ v(i,0)}{\sum_{j=1}^{n}\sum_{i=1}^{n} v\ (i,0)} \tag{9}$$

$$(j) = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} j\ v(i,0)}{\sum_{j=1}^{n}\sum_{i=1}^{n} v\ (i,0)} \tag{10}$$

Figure 2. Graphical representation of grid plotted burglary events depicted utilizing a white hot rendering to show higher concentration in comparison to the low concentration of events as depicted in purple.

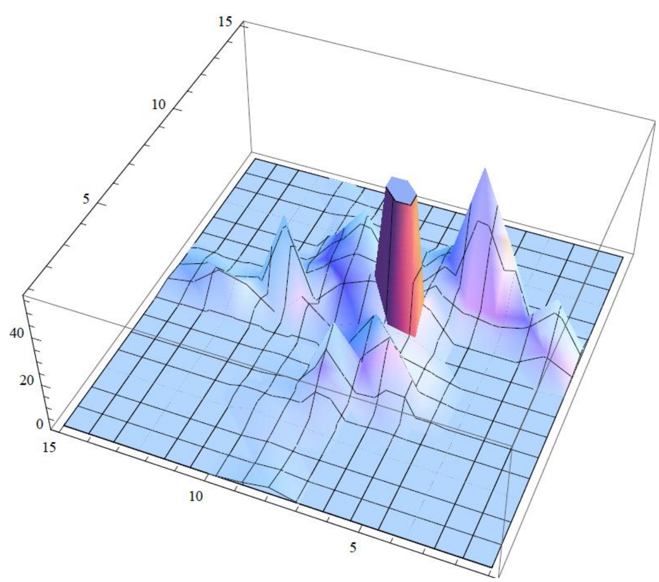


Figure 3. Graphical representation of epicenter where 9.10 units of x-axis and 7.30 units of y-axis where identified as the epicenter of all the burglary activity in a 3-D rendering.

Once all the information was collected and encoded to a specific data type, a binary logistic regression analysis was conducted to examine factors associated with the burglary activity related to the epicenter. This was conducted by comparing the events of the area of incremental one kilometer radii to the overall area until the model was determined to be statistically insignificant.

First a sequence of models were built and tested, starting from one kilometer radius with an increment of one kilometer until the model became insignificant. An overall model evaluation was conducted to determine if the model provided a better fit to an observed data set in comparison to a intercept-only model, or null model which serves as a baseline comparison due to the lack of predictors.

Table 3

Model of study vs. constant-only model

Radii comparison	< km Sample	> km Sample	α	Results
< 1 km to > 1 km and greater	23	568	3.207	$\chi 2$ (1, N = 591) = 28.247, $p < 0.001$
< 2 km to > 2 km and greater	132	459	1.246	$\chi 2$ (2, N = 591) = 15.996, $p < 0.001$
< 3 km to > 3 km and greater	283	308	0.085	$\chi 2$ (2, N = 591) = 77.017, $p < 0.001$

It was determined the results of one kilometer radius in comparison to two and greater kilometers and two kilometers in comparison to three and greater kilometers models were shown to be statistically significant given the use of the identified predictors. Three kilometers in comparison to four and greater kilometers was not significant given the use of the predictors; however predictors within the model showed to be significant and in turn making the model significant to discuss. The results are provided in Table 3.

Classification tests were conducted of the model of the given data comparisons to determine the ability of the model accurately predicting the placement of the burglary events as they relate to the radii they are associated. The results are depicted in Table 4. Results depicted that the model could accurately classify the burglary events in the comparison groups, however overall classification capability of the model diminished as the comparison sizes became close to each other demonstrating the predictors were becoming insignificant for overall classification of events as the radii increased.

Table 4
Classification by radii

Radii Comparison	< km Sample	> km Sample	<km (%)	to >km (%)	Overall (%)
< 1 km to > 1 km and greater	23	568	0	100	96.1
< 2 km to > 2 km and greater	132	459	0	100	77.7
< 3 km to > 3 km and greater	283	308	80.6	59.7	69.7

To determine the effects of the individual predictors, a statistical analysis was conducted to determine the significance of the individual regression coefficients. The regression coefficients, Wald statistics, and odds ratios for each of the five predictors were calculated and tested. The results are shown in Table 5. According to the Wald criterion only certain predictors associated per radii comparisons predicted the event of burglary activity. For the three comparison models, connecters were found significant at; 1 km $P[\chi^2(1)>25.286]<0.001$, 2 km $P[\chi^2(1)>3.905]=0.048$, and 3 km $P[\chi^2(1)>9.019]=0.003$. Barriers were found to be significant at 2 km and 3 km comparison models; $P[\chi^2(1)>7.656]=0.006$ and $P[\chi^2(1)>66.692]<0.001$. Through the three comparisons it was determined that either connectors or barriers were significant to warrant a response or a combination of the two.

Table 5

Significant predictors within the model

Radii Comparisons	Predictors	*B*	Wald	*df*	*p*	*OR*
< 1 km to > 1 km and greater	Connectors	2.162	25.286	1	0	8.691
< 2 km to > 2 km and greater	Connectors	0.33	3.905	1	0.048	1.391
	Barriers	-0.251	7.656	1	0.006	0.778
< 3 km to > 3 km and greater	Connectors	-0.404	9.019	1	0.003	0.668
	Barriers	-0.692	66.692	1	0	0.501

Furthermore, in order to test the sensitivity of small scale of 1 km radii to determine if there was a change from one to nine kilometers with one kilometer comparisons. A test of the full model with all five predictors against a constant-only model was statistically significant of comparisons from one kilometer to four kilometers and five to nine kilometers. The predictors, as a set, reliably distinguish between the different radii as they increase by one kilometer from the epicenter to the nine kilometer radius. Table 6 depicts the actual results per each kilometer comparison.

Table 6

Model of study vs. constant-only model

Radii comparison	<km Sample	>km Sample	α	Results
1 to 2 kilometers	23	109	1.556	X^2 (3, N = 132) = 24.166, p < 0.001
2 to 3 kilometers	109	151	0.326	X^2 (2, N = 260) = 11.466, p = 0.003
3 to 4 kilometers	151	82	-0.611	X^2 (1, N = 233) = 21.760, p< 0.001
4 to 5 kilometers	82	136	0.506	No Significance
5 to 6 kilometers	136	45	-1.106	X^2 (3, N = 181) = 64.269, p< 0.001
6 to 7 kilometers	45	9	-1.609	X^2 (1, N = 54) = 14.289, p< 0.001
7 to 8 kilometers	9	32	1.003	X^2 (2, N = 41) = 13.852, p< 0.001
8 to 9 kilometers	32	4	-2.015	X^2 (1, N = 34) = 7.266, p = 0.007

These comparisons indicate the set of predictors as a set reliably distinguish between activities of different radii from the identified epicenter of activity up to four kilometers.

Table 7 depicts the classification breakdown per radii. Classification was varied by kilometer comparison being relatively strong from one to four kilometers and five to nine kilometers with results ranging from 74% to 83%, except for the two to three kilometer comparison resulting in 53%. These classifications depict the ability of the models to correctly classify the events per the kilometer radii based on the identified predictors.

Table 7
Classification by radii

Radii Comparison	<km Sample	>km Sample	<km (%)	to >km (%)	Overall (%)
1 to 2 kilometers	23	109	13	96.3	81.8
2 to 3 kilometers	109	151	12.8	82.1	53.1
3 to 4 kilometers	151	82	91.4	43.9	74.7
4 to 5 kilometers	82	136	Could not be classified		
5 to 6 kilometers	136	45	88.2	62.2	81.8
6 to 7 kilometers	45	9	100	0	83.3
7 to 8 kilometers	9	32	93.3	0	82.4
8 to 9 kilometers	32	4	54.5	86.7	78

Table 8 shows the regression coefficients, Wald statistics, and odds ratios for each of the four predictors. According to the Wald criterion only certain predictors associated per radii comparisons predicted the event of burglary activity. Connector and barriers were significant predictors from one to three kilometers from the epicenter and from five to six kilometers. Repeat victimization was only significant within one to two kilometers and day of week from five to six kilometers.

Table 8

Significant predictors within the model

Radii Comparisons	Predictors	*B*	Wald	*df*	*p*	*OR*
1 to 2 kilometers	Repeat Victimization	1.481	4.534	1	0.033	4.396
	Connectors	2.509	16.234	1	0	12.296
	Barriers	0.895	10.624	1	0.001	2.448
2 to 3 kilometers	Connectors	0.567	8.739	1	0.003	1.763
	Barriers	0.361	6.006	1	0.014	1.435
3 to 4 kilometers	Barriers	-0.634	19.618	1	0	0.53
4 to 5 kilometers		No Significance				
5 to 6 kilometers	Day of the Week	-0.234	4.5	1	0.034	0.761
	Connectors	-3.31	10.24	1	0.001	0.037
	Barriers	-1.137	16.59	1	0	0.321
6 to 7 kilometers		No Significance				
7 to 8 kilometers		No Significance				
8 to 9 kilometers		No Significance				

Subsequent comparisons were conducted in addition to the procedure discussed earlier in the methods section to determine if there was any difference in the radii utilized for comparison. The first such comparison was observing the effects of the radii from one to nine kilometers in comparison to the epicenter or one kilometer radii, e.g. 1 km to 2 km, 1 km to 3 km, 1 km to 4 km. Comparison and results can be observed in Tables 9 to 11. As with the original procedure, the study model demonstrated a better fit utilizing the predictors, and the predictors significantly depicted between the different radii. However the results were similar to the one kilometer incremental increase with respect to the inconsistencies in the predictors in relation to the model and prediction of activity within the different radii.

Table 9

Model of study vs. constant-only model

Radii comparison	<km Sample	>km Sample	α	Results
1 to 2 kilometers	23	109	1.556	X^2 (3, N = 132) = 24.166, p< 0.001
1 to 3 kilometers	23	151	1.882	X^2 (2, N = 174) = 37.434, p< 0.001
1 to 4 kilometers	23	81	1.259	X^2 (1, N = 104) = 27.817, p< 0.000
1 to 5 kilometers	23	136	1.777	X^2 (1, N = 159) = 34.079, p< 0.001
1 to 6 kilometers	23	45	0.671	X^2 (3, N = 68) = 16.760, p< 0.000

Table 10

Classification by radii

Radii Comparison	<km Sample	>km Sample	<km (%)	to >km (%)	Overall (%)
1 to 2 kilometers	23	109	13	96.3	81.8
1 to 3 kilometers	23	151	39.1	97.4	89.7
1 to 4 kilometers	23	81	39.1	98.8	85.6
1 to 5 kilometers	23	136	39.1	100	91.2
1 to 6 kilometers	23	45	65.2	95.6	85.3

Table 11

Significant predictors within the model

Radii Comparisons	Predictors	*B*	Wald	*df*	*p*	*OR*
1 to 2 kilometers	Repeat Victimization	1.481	4.534	1	0.033	4.396
	Connectors	2.509	16.234	1	0	12.296
	Barriers	0.895	10.624	1	0.001	2.448
1 to 3 kilometers	Connectors	2.838	21.729	1	0	17.076
	Barriers	0.6	8.526	1	0.004	1.822
1 to 4 kilometers	Connectors	2.466	15.461	1	0	11.78
1 to 5 kilometers	Connectors	3.21	17.017	1	0	24.774
1 to 6 kilometers	Barriers	-0.933	13.308	1	0	0.393

Another analysis was conducted increasing the radius size and comparisons by two kilometers; two versus four kilometers and four to six kilometers, to utilize more data from the variables to reduce the variability. The comparison yielded the same results in terms of the model of study versus a constant-only model, use of the predictors would make the model significant. However this time it was determined only connectors were significant in the prediction of events between the different radii for both comparison and barriers specifically for the four to six kilometer comparison. Table 12 to 14 depicts the results of the two kilometer incremental comparison.

Table 12

Model of study vs. constant-only model

Radii comparison	<km Sample	>km Sample	α	Results
2 to 4 kilometers	132	233	0.568	X^2 (1, N = 365) = 14.885, $p< 0.001$
4 to 6 kilometers	233	181	-0.253	X^2 (2, N = 414) = 39.900, $p< 0.001$

Table 13

Classification by radii

Radii Comparison	<km Sample	>km Sample	<km (%)	to >km (%)	Overall (%)
2 to 4 kilometers	132	233	18.2	97.9	69
4 to 6 kilometers	233	181	79.4	46.4	65

Table 14

Significant predictors within the model

Radii Comparisons	Predictors	*B*	Wald	*df*	*p*	*OR*
2 to 4 kilometers	Connectors	0.586	13.276	1	0	1.797
4 to 6 kilometers	Connectors	-0.712	15.333	1	0	0.491
	Barriers	-0.485	26.207	1	0	0.616

Overall there are certain predictors; connectors, barriers, repeat victimization, and day of the week, that are required to depict between burglary activity from an identified epicenter in comparison to growing radii from that epicenter. However given the inconsistency of those predictors from a gradual increase of one to nine kilometers from that identified epicenter, the model would not be able to accurately predict the burglary activity given the predictors in relation to the distance from an identified epicenter. Based on this analysis by conducting a Somer's D analysis on the significance of the predicted probabilities it was determined the probabilities were insignificant and could not be utilized to predict the burglary activity for the different area comparisons.

Discussion

This study utilized logistic regression models in order to predict the probability of burglary activities with respect to the event density epicenter. The predictors of time of day, day of the week, connectors, barriers, and repeat victimization were selected based on prior studies that showed these variables to be more than sufficient to predict or correlate burglary activity. Results showed that certain variables such as connectors and barriers had significant effects when observed in certain radii size comparisons; however there was variation in the predictors of time of day, day of week, and repeat victimization. The model when utilized within all comparisons showed to be a usable model given the selected predictors, except for one radii comparison of four to five kilometers. The results of this study showed there was too much variation in the predictors of the given data to determine a specific pattern associated with burglary targeting or pattern prediction based on the distance comparisons utilized.

In this study the models for the initial radii size comparison of one kilometer to the remaining data population, two kilometers to the remaining data population, and the three kilometers to the remaining data population were determined to be significant models when utilizing the predictors of time of day, day of week, repeat victimization, connectors and barriers.

Only the three to remaining data population comparison was determined to be insignificant when the real world data was utilized to validate and verify the model. As shown with additional radii size comparisons; incremental increase in kilometers, incremental radii increase in relation to one kilometer and larger radii comparison models could be utilized to determine probability occurrence of burglary activity. However the results generated from the Daytona Beach area determine the application of the individual and the combination of the predictors to insignificant producing a wide range of probabilities which would not predict the occurrence of burglary activity.

Time of day and day of week were demonstrated by Ratcliffe (2004) to be characteristic of a burglar's decision making process, that burglars would utilize a specific time of day and day of the week that would best allow them to target a residence with the least likelihood of an encounter with the resident. For this study given the different demographic areas and other criminal activity that takes place, the burglaries in the Daytona Beach population area might be naturally random to determine a specific time of day or day of the week a burglar wishes to target. As the burglary events were utilized in the different radii comparison groups the variation between the data continued to become too great to determine any significance within the time of day and day of the week predictors. This shows the burglary events occurred at different times, never concentrating on any specific time or category.

Nee and Meenaghan (2006) conducted a study which interviewed 50 convicted burglars on their motives and reasons for conducting a burglary as well as the means to their decision making process. They determined that burglars had multiple reasons for committing an act; money, thrill, and to support a drug habit were the majority of means to commit the crime. However, they found that the individual would search out an area when the need to commit the crime had to satisfy one of the aforementioned reasons, which can be best described as stochastic. There would be no discernable pattern or set of circumstances that would allow the burglar to target a specific residence. This could be the case for the data sample utilized in this study, because the data was compiled from all burglary activities within the Daytona Beach area. In this study, the decisions of multiple burglars within a nine kilometer area are all different in a burglar’s target selection process.

For the factor of repeat victimization, Bernasco and Nieuwbeerta (2005) demonstrated that a burglar was most likely to be victimized again given the familiarity a burglar has with a residence and the ease they were able to target that residence originally. With this knowledge a burglar would target the same residence once more to allow additional time to search out higher pay off items with the perception the risk is now reduced given the additional knowledge of the residence. Pitcher and Johnson (2011) demonstrated repeat victimization would occur but only if there were “boosters” added to a predictive model to strengthen the relationship of the clustering of time and space with the added predictor of repeat victimization. They found that characteristics of residence, the distance from previous activity, as well as the relation of time in respects to repeat victimization all have to be considered when developing a model to predict burglary activity. In this study the level of detail associated with repeat victimization was not explored, and would strengthen the proposed model allowing for more precise prediction. A lot of events did fall in the same location if not within close proximity of the previous location. By examining these factors even closer, a more discernable effect could have been identified and modified to be incorporated directly into the model. Based on the layout of the Daytona Beach area, the repeat victimization was not the action of the same offender but of different burglars to generate an affect of repeat victimization. The addition of multiple offenders within a given area without prior knowledge and history of the targeted location contributed to repeat victimization to be significant only in the one to two kilometer radii comparisons.

For the factors of connectors and barriers, Peeters (2007) demonstrated the presence of connectors and barriers were directly related to a burglar's decision making process in relation to access to a multi-lane road system within a city. Access to this multi-lane system would allow an offender to immediately leave a targeted area whether or not law enforcement were contacted and dispatched, the offender would not be confined within a neighborhood and risk detection and apprehension. Barriers would also be considered as not to confine an offender to a particular area in the event the offense was detected. In this study it was found that connectors and barriers were significant in all comparison radii, but the effect was inconsistent in different radii sizes with respect to the identified epicenter. A similar study, Clare, Fernandez, and Morgan (2009) showed that connection to neighboring targeted areas in relation to the offender's residence was the defining characteristic for target selection, such as a mass transit system. Having a means to travel to the targeted area was influential to target selection as long as connectors negated the presence of the barriers associated with the targeted area. In the Clare et al. (2009) study the defining characteristic for connectors was a railway system and barriers were actual rivers. In this study access to a major road way were defined as connectors, and barriers restricted movement to those roadways similar to Peeters' study. The contradiction between the studies of the results concerning connectors and barriers can be attributed to the massive single lane road system that interconnects neighborhoods of Daytona Beach, essentially eliminating the need for the multi-lane road system for egress from a targeted location. The size of the neighborhoods in Daytona Beach also affords additional concealment by allowing a burglar to access different areas effectively expanding the area searched by law enforcement following a crime and reducing detection.

Hammond and Young (2011) demonstrated that crime will effectively reduce as it expands from the identified epicenter. In this particular study crime appeared to reduce as it expanded from the identified epicenter. However the identified epicenter was calculated based on all results over a given area. Based on observation the majority of the crime took place in a high demographic area that is characteristic of low economical and financial income which results in 15% of the population falling below the poverty level. This is greater than the national average as reported by the 2009 U.S. Census Bureau report. By incorporating all of Daytona Beach to facilitate the data, this inclusion eliminated significant effects that would normally be demonstrated by specific types of demographic, economics, and financial influences. Haddad and Moghadam (2010) showed that isolating a specific demographic and economics associated with a particular area would yield stronger results when determining the contributors associated with a particular crime, in this case burglary.

Additionally as demonstrated by Johnson, Bernasco, Bowers, Elffers, Ratcliffe, Rengert, and Townsley (2007) that clustering of events demonstrates a stronger relationship as it relates to time and geographic location. In this study predictor results became statistically insignificant as the radii size comparisons were increased. The Johnson et. al. (2007) demonstrated that crime would be associated in smaller areas essentially clustering in smaller groups. The size of the area must be taken into great consideration as not to influence the factors used to examine a specific area and diluting the effect of those predictors. As is the case with the Daytona Beach area, it can easily be seen which neighborhoods were more concentrated and clustered when observing a smaller area of comparison. By observing these smaller areas, the effect of predictors of time of day, day of week, connector, and barriers will become significant, which in turn would have strengthened the predictive model of this study.

Conclusion

The study used logistic regression to investigate the factors associated with burglary activity in Daytona Beach, Florida. It was found that these predictors of repeat victimization, connectors, and barriers were significant in determining a burglary pattern at certain radii comparisons. Through extensive analysis and further research it was found the predictors utilized are significant but the size of the observation area of Daytona Beach must be refined to strengthen the predictive model. The Daytona Beach Police Department would stand to benefit from this study by utilizing the model within a more constrained area observing the same factors as described in this study. Applying the procedures and model as discussed in this study within a one kilometer area with a 100 meter radii incremental comparison, will provide the resolution required for predictive analysis.

The use of the logistic regression model from this particular study will only be limited to identifying the probability of burglary activity. This will only provide information as it relates to the likelihood of burglary activity within the 100 meter incremental radii distances of the cluster's epicenter. This model will not identify specific locations, specifically residences, but will encompass the roads associated with an identified area in turn contributing to the police department patrol plan to deter or capture burglary suspects.

The application of a nonlinear regression model will potentially improve the strength of the study based on the parameters associated with conducting a nonlinear regression. Nonlinear regression observes the combination of physical and biological variables and quantitatively conceptualizes a process of interest. In this particular study this would be attributed to the behavior of the burglar in regards to time of day and day of week as it relates to connectors, barriers, and repeat victimization.

Events must be clustered more precisely within a one kilometer of each other, only then will the strength of the relationship of time and geographic location is significant. For example identifying relatively small neighborhoods with low incomes and substandard housing within a one kilometer area will increase the predictive occurrence by identifying a cluster based on demographics. Within the condensed area of observation, repeat victimization will become significant and generate a different effect as shown within this study. Further research into statistically significant demographics in relation to burglary will greatly increase the focus area of a study. This in turn will refine the association of connectors and barriers as it relates to a targeted residence.

Given the conclusion that smaller geographic areas are required for stronger time and space analysis, connectors and barriers potentially no longer will be influential given the observation the burglar will remain in a neighborhood they too utilize for residence. On the contrary, connectors and barriers can be further refined to determine the amount of association the targeted residence has with the actual residence of the described offender. Overall greater detail needs to be incorporated into a study that will utilize the predictors of time of day, day of the week, connectors, barriers, and repeat victimization. As demonstrated by this study encompassing a wide range of data over a large area with no refinement will yield insignificant results. This study can further be refined to observe smaller clusters of activity; this will allow law enforcement professionals to focus patrol efforts within a one kilometer residential area.

In reference to utilizing this study for other modes such as other law enforcement activities or IED pattern analysis and prediction the same approach can still be utilized when specific targeted areas of interest are identified based on activity and proximity to one another. The reduction of the observed area along with detailed analysis of the events within that given area will increase the significance of the predictors as well as develop an accurate prediction model.

References

A History of Crime Analysis. (2011). Denver, MA: Massachusetts Association of Crime Analysts. Retrieved from www.macrimeanalysts.com/articles/historyofcrimeanalysis.pdf

Alberto, P.A. & Troutman, A.C. (2003). *Applied behavior analysis for teachers*. Upper Saddle River, NJ: Merrill, Prentice-Hall.

Bernasco, W. & Nieuwbeerta, P. (2005). How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *Brit.J.Criminol,44*, 296-315.

Brower, M.C. & Price, B.H. (2001). Neuropsychiatry of frontal lobe dysfunction in violent and criminal behavior: A critical review. *Journal of Neurology, Neurosurgery and Psychiatry, 71*, 6.

Buck, G.A., Austin, R., Cooper, G., Gagnon, D., Hodges, J., Martensen, K., & O'Neal, M. (1973). *Police crime analysis unit handbook.* Washington, DC: US Department of Justice.

Canter, D. & Hammond, L. (2006). A comparison of the efficacy of different decay functions in geographical profiling for a sample of U.S. serial killers. *Journal of Investigative Psychology and Offender Profiling, 3*, 91-103.

Clare, J. Fernandez, J., & Morgan, F. (2009). Formal evaluation of the impact of barriers and connectors on residential burglars' macro-level offending location choices. *The Australian and New Zealand Journal of Criminology, 42*, 139-158.

Cox, D. R. & Snell, E. J. (1989). *The Analysis of Binary Data*. London: Chapman and Hall.

Curtin, K. M. (2011). Linear referencing for network analysis of IED events. Retrieved from George Mason University Department of Geography and GeoInformation Science.

Eck, E., Chariney, S., Cameron, J., Leitner, M., and Wilson, R. (2005). *Mapping crime: Understanding hot spots.* Washington, DC: US Department of Justice.

Elffers, H. (2004). Decision models underlying the journey to crime. *Punishment, Places, and Perpetrators: Developments in Criminology and Criminal Justice Research*, 182-197, Willan Publishing, Uffculme Cullompton.

Ferrari, S., Baumgartner, K. C., Palermo, G. B., Bruzzone, R., & Strano, M. (2008). Network models of criminal behavior: Comparing Bayesian and neural networks for decision support in criminal investigations. *IEEE Control Systems Magazine*.

Fritzon, K. (2001). An examination of the relationship between distance travelled and motivational aspects of firesetting behavior. *J Environ Psychol, 21*, 45-60.Gottfredson,

M. & Hirschi, T. (1990). *A General Theory of Crime*. Stanford, CA: Stanford University Press.

Greenberg, S.W. and Rohe, W.M. (1984). Neighborhood design and crime: A test of two perspectives. *Journal of the American Planning Association*, 50(1), 48-60.

Goldstein, H. (1990). *Problem-oriented policing*. New York, NY: McGraw-Hill.

Haddad, G.R.K. & Moghadam, H.M. (2011). The socioeconomic and demographic determinants of crime in Iran (a regional panel study). *Eur J Law Econ*, *32*, 99-114.Harris, K.D. (1974). *The Geography of Crime and Justice*. McGraw Hill, New York.

Hammond, L. & Youngs, D. (2011). Decay functions and criminal spatial processes: Geographical offender profiling of volume crime. *Journal of Investigative Psychology and Offender Profiling*, 8, 90-102.

Hawkins, David, J., Guo, J., Hill, K.G., Battin-Pearson, S., & Abbott, R.D. (2001). Long-term effects of the seattle social development intervention on school bonding trajectories. *Applied Developmental Science*, *5*, 225-236.

Horney, J. (2006). An alternate psychology of criminal behavior. *The American Society of Criminology 2005 Presidential Address*, *44*, 1.

Hosmer, D. W. & Lemeshow, S. (2011). *Applied logistic regression*. New York, New York: Wiley-Interscience Publication.

Linear Regression. (2011). Retrieved from http://www.stat.yale.edu/Courses/1997-98/101/linreg.html.

Jing, W. & Tao, Z. (2011). Analysis of decision tree classification algorithm based on attribute reduction and application in criminal behavior. *IEEE*.

Johnson, S.D., Bernasco, W., Bowers, K.J., Elffers, H., Ratcliffe, J., Rengert, G., & Townsley, M. (2007). Space-time patterns of risk: A cross national assessment of residential burglary victimization. *J Quant Criminol, 23*, 201-219.

Kai, Y.T., Heng, M.A., Bullock, L.M. (2007). What provokes young people to get into trouble: Singapore stories. *Preventing School Failure, 51*, 2.

Kleinbaum, D.G., Klein, M., & Pryor, E.R. (2010). Logistic Regression: A Self-Learning Text. Statistics for Biology Health ed. 3. Springer.

Krandel, E. & Freed, D. (1989). Frontal-lobe dysfunction and antisocial behavior: a review. *F Clin Psychol, 45*, 404-413.

Laub, J. (2005). *Edwin H. Sutherland and the Michael-Alder Report: Searching for the Soul of Criminology 70 Years Later*. Edwin H. Sutherland Address, American Society of Criminology. Toronto, Ontario.

Law, D. M., Shapka, J. D., & Olson, B. F. (2010). To control or not to control? Parenting behaviours and adolescent online aggression. *Computers in Human Behavior, 26*, 1651-1656.

Mason, S.J. & Baddour, O. (2007). Chapter 7 Statistical modeling. In A. Troccoli, M. Harrison, D. L. T. Anderson, & S. J. Mason (Eds.), *Seasonal Climate: Forecasting and Managing Risk* (pp.167-206). Dordrecht, The Netherlands: Springer.

Menard, S. (2002). *Applied logistic regression*. London, England: Sage Publications.

Mischel, W. & Shoda, Y. (1995). A cognitive-affect system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246-268.

Nagelkerke, N.J.D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78, 691-692.

Nash, D. C., Martin, T. L., Ha, D. S., & Hsiao, M. S. (2005). Towards an intrusion detection system for battery exhaustion attacks on mobile computing devices. *Proceedings of the 3rd International Conference on Pervasive Computing and Communications Workshops*.

Nee, C. & Meenaghan, A. (2006). Expert decision making in burglars. *Brit.J.Criminol, 46*, 935-949.

Peeters, M.P. (2007). The influence of physical barriers on the journey-to-crime of offenders. *The Netherlands: Netherlands Institute for the Study of Crime and Law Enforcement*.

Peng, C.J., Lee, K.L., & Ingersoll, G.M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96, 1-14.

Pillai, G., & Kumar, R. (2007). Simulation of human criminal behavior using clustering algorithm. *International Conference on Computational Intelligence and Multimedia Applications*.

Pitcher, A., & Johnson, S. (2011). Exploring theories of victimization using a mathematical model of burglary. *Journal of Research in Crime and Delinquency,* 48, 83-109.

Ratcliffe, J.H. (2004). The hotspot matrix: a framework for the spatio-temporal targeting of crime reduction. *Pol Prac Res, 5*, 5-23.

Ratcliffe, J.H. & McCullagh. (1998). The perception of crime hot spots: a spatial study in Nottingham, U.K. *Crime mapping case studies: success in the field*, 45-51. Police Executive Research Forum, Washington, DC.

Rattner, A., & Portnov, B., (2007). Distance decay function in criminal behavior: a case of Isreal. *Ann Reg Sci*, 41, 673-688.

Rengart, G.F. (2004). The journey to crime. *Punishment, Places, and Perpetrators: Developments in Criminology and Criminal Justice Research*, 169-181, Willan Publishing, Uffculme Cullompton.

Sagovsky, A., & Johnson, S.D. (2007). When does repeat burglary victimization occur? *The Australian and New Zealand Journal of Criminology, 40*, 1-26.

Sherrod, P.H. (2011). *DTREG Predictive Modeling and Forecasting*.

Short, M.B., D'Orsogna, M.R., Brantingham, P.J., & Tita, G.E. (2009). Measuring and modeling repeat and near-repeat burglary effects. *J Quant Criminol, 25*, 325-339.

Snook, B., Dhami, M.K., & Kavanagh, J.M. (2011). Simply criminal: Predicting burglars' occupancy decisions with a simple heuristic. *Law Hum Behav, 35*, 316-326.

Sterzer, P. (2010). Born to be criminal? What to make of early biological risk factors for criminal behavior. *The American Journal of Psychiatry, 167*, 1.

Swaim, R.C., Kimberly, L.H., & Kathleen, K. (2006). Predictors of aggressive behaviors among rural middle school youth. *Journal of Primary Prevention, 27*, 229-243.

Townsley, M., Homel, R., & Chaseling, J. (2003). Infectious burglaries: A test of the near repeat hypothesis. *Brit.J.Criminol, 43*, 615-633.

Trejos-Castillo, E., Vazsonyi, A., & Jenkins, D.D. (2008). Violent and criminal behavior in rural and non-rural African American youth: A risk-protective factor approach. *Southern Rural Sociology, 22*, 108-130.

Trout, A.T., Nordness, P.O., Pierce, C.D., & Epstein, M.H. (2003). Research on the academic status of children and youth with emotional and behavioral disorders: A review of the literature from 1961-2000. *Journal of Emotional and Behavioral Disorders,11*, 198-210.

Weisberg, S. (2005). *Applied Linear Regression*. New York, NY: Wiley-Interscience.

Weisz, A. N., Tolman, R. M., & Saunders, D. G. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *J Interpers Violence, 15*, 75-90.

Zabel, R.H. & Nigro, F.A. (2001). The influence of special education experience and gender of juvenile offenders and academic achievement scores in reading, language, and mathematics. *Behaviorial Disorders*, *26*, 164-172.