Predicting the Primary Residence of Serial Sexual Offenders: Another Look at a Predictive Algorithm

Jenelle Louise (Taylor) Hudok[1,2]
[1]*Department of Resource Analysis, Saint Mary's University of Minnesota, Minneapolis, Minnesota 55404*, [2]*Minnesota Bureau of Criminal Apprehension, Saint Paul, Minnesota 55106*

Keywords: predatory offender, sex(ual) offender, Journey-to-Crime, CrimeStat, GIS, crime, geographic profiling

Abstract

This study analyzes a series of crimes committed by sexual offenders and a series committed by non-sexual offenders. Subjects chosen for the study were seven sexual offenders (Test Group) and twelve offenders with non-sexual crimes (Control Group). Test data were selected from the Minnesota Predatory Offender Registry. Control group data were randomly selected from data provided with the crime analysis software used in this study (CrimeStat III). Historical offense data from all offenders were analyzed using the Journey-to-Crime (JTC) function of CrimeStat III to predict the most likely location of residence based upon the location of offenses. The distance predicted was compared with the actual location of the residence of each offender. Data collected for each group was compared both visually and statistically to determine if there was a significant difference between the two groups.

Introduction

In the 1980's a trend in crime analysis called geographic profiling became popular. Computerized mapping of crime incidents began to replace and improve upon "pin maps" which had been utilized for decades to analyze geographic trends in crime (Rich and Shively, 2004). Recent studies into crime data have led to the development of computer software that can be used to predict the residence of a serial offender by looking at the pattern of offenses committed. Many of these software packages require a series of at least five crimes in order to be effective (Levine, 2004). The purpose of this study is to determine to what extent an algorithm provided in one of these software packages can be used to determine a pattern in crimes of a sexual nature committed by the same offender with fewer than the recommended number of incidents.

This study seeks to show that because of the particular nature of sexually motivated crimes, fewer incidents are needed to accurately estimate the offender's origin. To test this, data on serial sex offenders was compared to control data of offenders who committed non-sexual crimes.

St. Paul Cases – When is it enough?

In the early morning hours of New Year's Day, 2007, a brutal stranger rape occurred in the City of Saint Paul, Minnesota. The Saint Paul Police Department began to investigate the case immediately, a suspect description was

Hudok, Jenelle Louise (Taylor). 2008. Predicting the Primary Residence of Serial Sexual Offenders: Another Look at a Predictive Algorithm. Volume 10, Papers in Resource Analysis. 10pp. Saint Mary's University of Minnesota Central Services Press. Winona, MN. Retrieved (date) from http://www.gis.smumn.edu.

compiled and evidence was gathered. One week later, on January 7th, 2007, a second stranger rape occurred only a few blocks from the previous incident. When St. Paul Police realized that the suspect description provided by witnesses in the second incident was strikingly similar to the description from the incident one week prior, they became concerned that they had a serial rapist in their city (Volpe, 2007). Though these offenses were deemed to be unrelated as a result of DNA analysis months after they occurred (Croman, 2007), the police department's initial concerns beg some tough questions.

A series of crimes can be composed of as few as two offenses. Yet, as stated previously, many models used to predict the residence of an offender based on the series of offenses require five incidents or more in order to be effective. Is this helpful in the case of a serial rapist who is just beginning a crime spree? Do the police need to wait for three more victims before they can begin utilizing these models to attempt to pinpoint a possible location of the offender? Is it possible to effectively utilize these models with fewer than five incidents?

Background

As of 1995, every state has established a "registry" of persons who have committed sexual offenses (Matson and Lieb, 1996). In 2007 the federal government greatly expanded registration on a nation-wide scale when it passed the Adam Walsh Act. This legislation requires a broader scope of all state registries and establishes a more centralized federal registration initiative. According to Matson and Lieb, one of the cited reasons for registration laws is to create "a tool that law enforcement uses to solve crimes or, ideally, to prevent them" (Matson and Lieb). The information stored by registries consists of the offender's name, address, work place and data on vehicles driven by the offender (Matson and Lieb). This information is stored primarily to serve the purpose of finding these offenders in the event of an incident and secondarily to try to prevent these offenders from committing future crimes by providing information to members of the general public. The use of registration as a preventive measure is designed to make the public aware of the location and history of sex offenders, thus making them a participant in the safety of their own community.

Communities have a tendency to react quickly and decisively when an incident of criminal sexual conduct is committed in their midst. Members of the public want perpetrators of these crimes identified and caught very quickly. Both the data that is collected and current government practices vis-à-vis registration indicate that offenders who commit sex crimes tend to be a different sort of criminal than offenders who commit other types of offenses (Griffin and West, 2006).

The difference, according to Blake and Gannon (2008), is largely in the *modus operandi* (method of operations) of the offenders. In discussing treatment for sex offenders, Blake and Gannon indicate that sexual offenders have differing thought processes than non-sexual offenders. Some of these differences are rooted in "cognitive distortions, social perception, and empathy" (Blake and Gannon). Many treatment programs for sex offenders focus on correction of theses psychological deviations. No other type of crime has such advanced and scrutinized treatment options. There are many treatment programs for addictions, but research discovered no treatment programs for a burglar, a robber or an arsonist.

These psychological differences between sexual offenders and non-sexual offenders may play a role in the geographic pattern of crimes committed by these offenders. The proposed theory for this study posits that sex offenders are more predictable in their crimes because of these psychological factors. If these offenders are more predictable, it would follow that fewer than the recommended number of incidents is needed to determine a pattern. Therefore the algorithm's prediction of the offender's primary residence will be closer to its actual location. This is the theory that will be tested in this study.

Methods

The Software

In the mid-1990s, a crime analysis package called CrimeStat was designed and provided free-of-charge for students and criminal justice agencies (Levine, 2004). The current version of this program is CrimeStat III. This software package provides a function called Journey-to-Crime (JTC) analysis. This function utilizes an algorithm by the same name to evaluate a series of crimes and uses that data to predict the primary residence of the perpetrator of those crimes. In order to be accurate, Levine states that the algorithm requires a series of at least five crimes in order to recognize a pattern and thus a possible location of the perpetrator.

Data Collection

Data for this project were collected and analyzed in several steps. The first step involved choosing subjects for the study from the Minnesota Sex Offender Registry housed at the Minnesota Bureau of Criminal Apprehension in Saint Paul, Minnesota. The second step involved collecting data on the offenses committed by each offender and geocoding the locations of each. Thirdly a Journey-To-Crime analysis was performed on the data. Then control data was retrieved to compare to the sex offender data. This control data were extracted from data provided with the CrimeStat III software package. Finally, these data were analyzed using the same method as the sex offender data and the two results were compared.

Collecting Sex Offender Data

The search for subjects for this study began with compiling a list of criteria for the study group. The defined requirements were:

- Subject must have a Level 2 or Level 3 designation. Certain information on Level 2 and Level 3 offenders who live in the community is deemed to be public information by the State of Minnesota. Information on other sex offenders is confidential in Minnesota and is for law enforcement use only.
- Subject must have committed multiple offenses in order to be included. Offenders with only one offense would not contribute to this study because they have committed no series of crimes to analyze.
- The offense must be an incident involving a stranger. That is,

victims must be designated as unknown to the offender. The purpose of this is to include only offenses that occur out of the home for this study. Many offenses that involve victims previously known to the offender are intra-familial and occur in the home.

- Victims of the offenses must be over the age of majority. This criterion was included to eliminate the possibility of using confidential data on minors. Though every attempt is made to eliminate identifying data, it is important to exclude these victims.

A list of 120 offenders meeting these criteria was retrieved from the Minnesota Predatory Offender Registry. Eight offenders were randomly selected from this list and again reviewed in more depth for these same criteria. All identification information on the offenders was deleted out of the working file created for this study and identification numbers were assigned to each offender randomly.

Note that although seven offenders were chosen, there are eight series that were analyzed in this study. One of the selected offenders had two very distinct series of crimes in two separate areas of Minnesota. That offender, therefore, accounts for two series of data presented in this study. Note also that it appears as if two separate groups are being included in this study, Level 2 offenders and Level 3 offenders. For the purpose of this study, there is essentially no difference between these two groups because their current risk of re-offending in the future, indicated by their assigned Risk Level, does not have an effect on crimes they have committed in the past. Since only historical crimes are used in this study, current re-offense probabilities are not relevant.

Information on the location of offenses committed by these offenders was then added to this data. This information was entirely in address format. The addresses were geocoded using ArcView 9.2 and saved into a new file. The lowest matching rate encountered during geocoding for addresses was 64% match for one address. This address was verified individually by the author and accepted as correct. Address data was replaced with x,y coordinates of each offense in meters using the North American Datum of 1983 and projected as Universal Transverse Mercator in Zone 15N. The same procedure was used on the home locations for each offender.

Control Data

Control data was retrieved from a set of data packaged with the CrimeStat III software. Twelve individuals were randomly chosen from this large dataset. Each of these offenders had one stationary home residence when they committed their series. The crime series in this dataset were committed in and around the area of Baltimore, Maryland. Each crime and home location was stored as x,y coordinates in latitude/longitude format. These values were projected into the North American Datum of 1983 and projected as Universal Transverse Mercator in Zone 18N. Two additional x,y fields were then added and populated with the values as projected in meters. The data then matched the data used for the sex offender group.

__Journey-To-Crime Analysis__

The Journey-To-Crime (JTC) function of the CrimeStat III software package is designed to be an easy-to-use windows-based interface (Levine, 2004) as shown in Figure 1.

Figure 1. CrimeStatIII Interface.

The data is not loaded into the software; rather, the user indicates which file to analyze and sets parameters for that analysis. Results can be stored in various formats including the one used in this study, the Environmental Systems Research Institute's shapefile format.

A test area must be defined for each analysis. This test area is an area in which the algorithm will calculate the probability of home locations. This test area is broken down into a grid for the purpose of the equation. For this study, the test area was created by drawing a large rectangle around the city or cities in which each offender's incidents and home are located. The boundaries were stored in CrimeStat III and utilized for this analysis. During the calculations, CrimeStatIII breaks this test area down into a grid and calculates the probability that each cell grid is the one in which the offender's residence is located. This study used a grid broken into 200 columns across the test area. This limitation on the test area prevents excessive processing time and less resulting data as a probability is calculated for each individual pixel of the resultant grid (Levine, 2004).

The data results from the JTC function are two different pieces of information that are important for different reasons. First, each pixel in the test area grid is assigned a probability that the offender lives in that grid, as stated previously. By mapping these probabilities and sorting into groups, a picture can be shown of the areas of most probability and concentric circles of less probability (Figure 2).

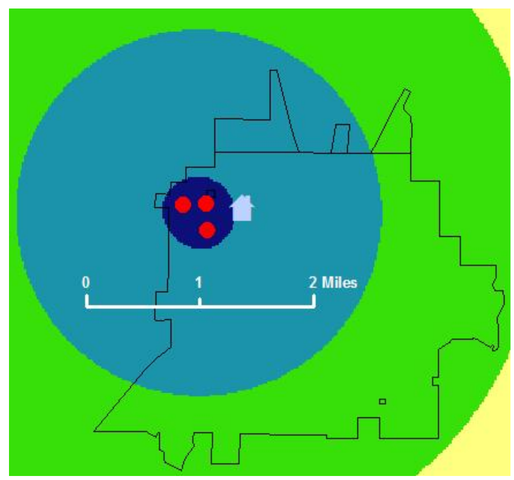

Figure 2. Map for Offender ID25. Showing the probability circles of the resultant maps.

Second, the JTC returns one point of peak likelihood. This point is an x,y coordinate in meters and was used in the statistical analysis of the data. This point is generally located in the center of the area designated as the circular point of most likelihood in the maps.

The same steps were taken with the test data provided with the CrimeStat III software package. Analysis was performed in the same manner with the same parameters set forth previously.

Analysis of Distances

As was previously stated, the resultant peak likelihood points for each sex offender were collected and the distance between the calculated peak likelihood point and the actual home residence of the offender was calculated. This information was compared to the same calculated information for the control subjects. The distance for each offender's series is shown in Table 1 and Table 2 below.

Table 1. Distance in miles of the point of highest probability to the actual residence of the sex offender.

Offender Identifier	Distance in Miles
ID06	0.265775
ID07	1.168591
ID17R	2.953385
ID17M	0.982244
ID20	6.861665
ID21	7.401207
ID25	0.357813
ID27	0.798954

Table 2. Distance in miles of the point of highest probability to the actual residence of the other offender.

Offender Identifier	Distance in Miles
S15A	0.541688
S13B	1.757405
S20A	11.586812
S12B	0.652204
S6B	3.017610
TS16B	0.020111
TS15A	1.886535
TS13F	10.332147
S6A	5.160217
S8C	2.415490
S16A	2.793044
S3C	0.677447

The data in these tables were then compared using an independent two-sample t-test to determine if the hypothesis was correct.

Results

The results of the analysis performed on these data were analyzed using two methods. First, maps depicting concentric rings of probability were created for each series of data. Analysis of these maps is purely visual. The maps present a very good indicator of the accuracy of the predictions on the part of the algorithm. Secondly, a statistical test was performed on the distances shown previously in Table 1 and Table 2.

Maps

Figure 3 shows a typical map created with the results of the analysis of each crime series. Each crime location is indicated by a dot and the offender's actual primary home location is indicated by a house shaped symbol.

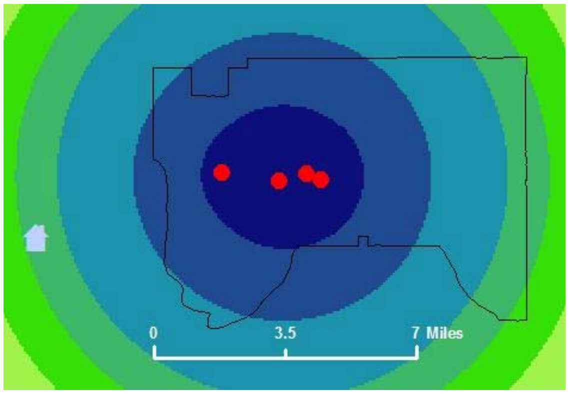

Figure 3. Map for Offender ID21. The residence is far outside of the area of highest probability.

Figure 3 shows the concentric rings of probability as discussed previously. In this circumstance, the offender lives in the third circle out from the center circle, the area of highest probability. These figures can provide assistance if a serial offender is suspected, however, as this figure demonstrates, other tools may be necessary as well. They can provide a guide to law enforcement to use while assigning officers to saturation details or arranging sweeps of offenders in a particular area, but should not be the only tool utilized in an investigation.

When viewing these maps, it is important to note the varying sizes of the areas of highest probability for each offender. Presumably, if the areas of highest probability are fairly small, different techniques can be used than if that area is very large. The more dispersed the offenses, the larger the area of highest probability for the offender's residence as seen in the figure for the offender identified as ID27 (Figure 4). The area of highest probability for this offender covers nearly half of the entire city of Saint Paul, Minnesota.

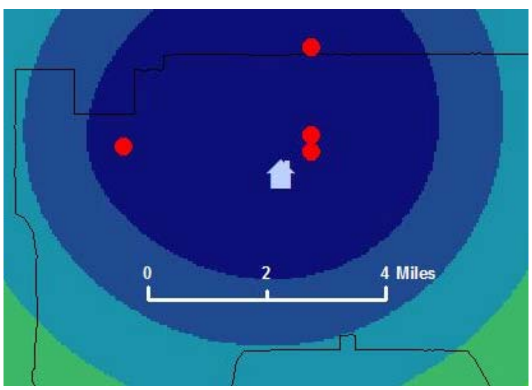

Figure 4. Map for Offender ID27. Note the large area of high probability and the dispersal of the offenses.

Not all residences for offenders used in this study fall within the area of highest probability calculated by their offenses. The figure for the offender identified as ID21, for example, shows an actual residence some distance outside of the area of highest probability (Figure 3). In this instance, the search area for an investigation could be falsely limited. This offender may not be found in the area since the person lives in an entirely different city.

In situations where the offenses are so dispersed, it is unlikely that an investigative team will feel the need to consult a map or an algorithmic analysis to limit their search area. Unless there is a pattern of geographically close incidents, these maps can be of little use. In the case of Offender ID27 (Figure 4), the distance from the point of highest probability to his actual residence is less than a mile. In looking at the map, however, uncertainty imbues the case because of the large area covered by the centermost circle. This is a good demonstration of the reason why, even were this algorithm more studied, law enforcement would utilize multiple tools in addition while working on these types of cases.

These difficulties in utilizing maps for truly analytical analysis prevent them from being overly useful by themselves. For this reason the author utilized other means to calculate the potential validity of the theory being tested in this study. Since these maps can at times give rise to more questions than answers, statistical analysis was employed to further clarify differences, or lack thereof, between the studied groups.

Statistics

The t-test statistic was performed on the data shown in Table 1 and Table 2. This test is designed to determine if the difference between the groups is statistically significant. For the purposes of the t-test, the test (alternate) hypothesis is: the distance between the predicted location and the actual location of the non-sexual offenders will be greater than the distance calculated for the sex offenders. That is, this tests the hypothesis that the predicted point of the location of the offender will be more accurate in the sex offender group than in the group of other offenders.

The calculated t-value was 0.749917 ($p > 0.05$, one-tailed test). Here, the t-test shows that there was not a statistically significant difference between the two groups of data used in this study.

Discussion and Conclusions

The statistical results of this study show, based upon data utilized here, that the originally stated theory was not supported. Visual analysis of the maps shows that, in some circumstances, the algorithm does give a limited area in which to begin looking for a serial offender. That area tends to be fairly accurate in most of these cases, but there are others where the offender is clearly a large distance outside of the area of highest probability. Offenders with this pattern will also have a large distance used in the aforementioned tables 1 and 2. The offender identified as ID20, for example, was predicted to live approximately 6.8 miles away from the actual residence (Table 1). It should be noted here that while the algorithm used predicts home residences, the possibility exists that the offender is also familiar and moves about in other areas farther from home. While not included in this study, further investigation of the ramifications of these phenomena is recommended. This may help to answer questions in regards to whether the algorithm only predicts home residence or whether it actually predicts the location of an address the offender frequents but where he does not live such as a workplace or a family member's residence.

Based upon these results, the utilization of this algorithm is not advised in an investigation if there are fewer than the five incidents in a series. With further study in this area, however, refining of the algorithm may be possible to account for fewer incidents or for specific types of offenders (i.e. sex offenders, drug offenders, property crime offenders, etc.).

The scale of this study is very small. Minnesota has 20,000 registered sex offenders and of those, only 120 fit the criteria for inclusion in this study. Of the 120, many have incomplete data and some have no supporting court documents from which a location of the crime could be ascertained. Also not included in this study are offenders who will never register because they will be incarcerated for the rest of their natural lives. These offenders will not register because registration only occurs upon the initial release from prison. Therefore, there is an entire pool of offenders in Minnesota that remains to be analyzed for such a study. It stands to reason that some of Minnesota's worst serial sexual offenders will remain in custody for the remainder of their lives. These offenders will be analyzed in a follow-up to this study conducted at the Minnesota Bureau of Apprehension for the Predatory Offender Investigations Unit. That study will be conducted in a manner similar to this study.

Limitations

There are several notable limitations to this study. First, as mentioned previously, the study was performed on a very small scale. A further limitation of this study was the location of the crime series of the control data. Data collected on the sexual offenders was all located within the state of Minnesota but the control data crime series were all located within the state of Maryland in the area of Baltimore. It is not known if the geographic difference affected the outcome of this study, but it is a consideration that should be accounted for in future attempts to approach similar research topics.

In addition, there are several additional studies that may be necessary in order to truly calculate the usability of this algorithm in a real world setting. A study into whether this algorithm may indicate residences or locations other than a home residence may need to be investigated. This research also relies heavily upon the algorithm's creator's assertion that there need be five incidents in order for the algorithm to function properly. This study did not test that assertion by analyzing series with five or more incidents. This would also be a follow-up study that may lend credence to this algorithm and its potential for use in a real world situation.

Acknowledgements

I would like to thank the Minnesota BCA, especially the Predatory Offender Registration and Predatory Offender Investigation Units for allowing me use of data in the Minnesota Predatory Offender Registry and for their support in the development and completion of this project. I would like to thank my undergraduate professors of geography at the University of Saint Thomas, Dr. Paul Lorah and Dr. Robert J. Werner for their nurturing of my early interest in GIS and crime mapping. And finally, I would like to thank in a very special way Dr. David McConville and Mr. John Ebert and everyone in the Resource Analysis Department for supporting and encouraging me throughout my years at Saint Mary's University of Minnesota.

References

Rich, T and Shively, M. 2004. A Methodology for Evaluating Geographic Profiling Software. Abt Associates, Inc., Cambridge, MA, and the National Institute of Justice, Washington, D.C. http://www.ojp.usdoj.gov/nij/maps/gp.html. (Retrieved on 24 August 2007).

Blake, E. and Gannon, T. 2008. Social Perception Deficits, Cognitive Distortions, and Empathy Deficits in Sex Offenders: A Brief Review. Trauma, Violence, & Abuse. 9:34-55.

Croman, J. 2007. Saint Paul teen charged in Eastside rape. Kare11 News website and print article. Retrieved 5 March 2007.

Griffin, M. P. and West, D. A. 2006. The Lowest of the Low? Addressing the Disparity between Community View, Public Policy, and Treatment Effectiveness for Sex Offenders. Law & Psychology Review. 30: 143-169.

Levine, N. 2004. CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington D.C. (Retrieved on 21 September 2007).

Matson, S and Lieb, R. 1996. Sex Offender Registration: A Review of State Laws. Washington State Institute for Public Policy.

Volpe, T. 2007. St. Paul Police investigate new leads in search for serial rapist. Kare11 News website news article. Retrieved 16 January 2007.