IDENTIFYING REPEAT VICTIMIZATION WITH GIS

J. H. RATCLIFFE and M. J. McCULLAGH*

Until recently most attempts at identifying repeat victimization locations have focused on searching address fields in police records. Problems with inaccurate data entry and variation in address format make this method fraught with difficulty and time consuming to correct. This study of burglary suggests that a standard GIS package, searching geo-referenced crime locations can dramatically improve the time and accuracy of identifying repeats. The research presented here appears generally consistent with other published work in that the period of highest risk is immediately after an initial burglary. The study covers a longer period than usual. Two years’ worth of data raises issues regarding definitions of repeat victimization.

The benefit to crime prevention of identifying repeat victimization has been widely recognized (Anderson et al. 1995; Ellingworth et al. 1995; Farrell and Pease 1993), but the process of accurately distinguishing the repeat locations has always been difficult. While the under-reporting of crime to the police is also well documented (Hough and Lewis 1989; Mayhew et al. 1993; Tilley 1995), it remains a reality that police recorded crime data are still one of the best sources of information on local crime distribution in this country. Computerized systems for recording police crime data have been set up within forces but usually the extraction of data pertinent to the geographical crime distribution and the identification of repeat victims is not a priority.

A number of articles highlight the difficulties posed by police data in identification of repeat victims (Read and Oldfield 1995; Sampson and Phillips 1995). Crime data tend to be recorded for statistical measurement, and not specifically designed for the identification of repeat victimizations (Ellingworth et al. 1995). A graphic example of the difficulties posed by police crime recording systems is provided by David Anderson (1995). The study was geographical only in confining the search parameters to a certain police divisional area, and the identification of repeat victims to home addresses.

These examples all show the problems of trying to identify a unique location from a number of text fields within a crime data system. If search criteria are tied to beat boundaries and other forms of containment then the problems of identifying repeats increase. Address complexity increases the geo-referencing problem. For example, Nottinghamshire Constabulary’s Crime Recording Interim System (CRIS) has separate fields for building name, building number, floor number, sub unit number, sub unit name, sub street name, street name. With such a variety of options it is hardly surprising that different operators on different shifts occasionally record a complex address in fields different from those of their colleagues. Nottinghamshire Constabulary has attempted to get round this problem by improving the address field data entry dialog on CRIS. The system will only accept addresses that it recognizes, and then places the

* Geography Department, University of Nottingham. The authors would like to thank Nottinghamshire Constabulary for access to the data and co-operation in this research. They would also like to express their appreciation to the reviewers of this manuscript for their helpful comments and advice.
correct house number, street and postcode in the right fields. This system can still fail as it
does not yet automatically enter flat numbers in shared buildings, or business addresses
in small industrial estates that change regularly. Nevertheless this is a considerable
improvement to a ‘free’ text entry system. Nottinghamshire Constabulary have gone to
considerable effort and expense to improve CRIS, an option not available to all forces.

The problem, therefore, is to find a method of extracting repeat victimization records
accurately and quickly from a crime database which is known to have conflicts within the
textual address fields. Accuracy is necessary to prevent a follow-up manual trawl through
the data, and speed is important because the highest chance of a repeat is in the first few
weeks after an initial incident. On a practical level, these factors are also important,
because if police forces are to utilize the information it must be seen to be worth the
investment in time and effort.

One possible solution is to utilize the spatial nature of geo-referenced data. This is
now possible as many police forces have access to geographically referenced data
through their growing use of Geographical Information Systems (GIS). GIS have
significant advantages to standard databases as ‘the key features which differentiate GIS
from other information systems are the general focus on spatial entities and
relationships’ (Maguire 1991). Police forces have been relatively quick to recognize the
advantages of using GIS to aid their deployment of resources, and the growth of GIS
within forces has been rapid (Campbell 1992; Clegg and Robson 1995; ICL 1995). Most
of the use of GIS within forces has been directed at resource allocation and a clearer
identification of the location of a 999 call. However some forces (including
Nottinghamshire) have been geo-referencing the crime incident records by including
the National Grid coordinates of the event on their systems. In the case of
Nottinghamshire, they began by tying in the location of the crime incident to the
Postcode Address File (PAF) which has a spatial resolution of 100 metres for the full
Postcode centroid. A number of interesting works comment on the accuracy and
usefulness (or not) of the PAF (Gatrell 1989; Raper et al. 1992). Recently the
Nottinghamshire CRIS system has been updated to take advantage of Address-Point data,
a commercial package available from the Ordnance Survey (OS) identifying each
property in the country with a unique (supposedly) 0.1 metre resolution National Grid
reference which is updated biannually. The CRIS upgrade has coincided with the more
rigid data entry dialog on the CRIS system considered earlier. As long as the central
database of properties, addresses and geo-references in the county is maintained, there
should be few locations which are not known to the system. The CRIS system
automatically assigns the grid location which corresponds to the address or location
chosen by the operator.

The inclusion of a geographical reference in the recorded crime data allows for a new
method of searching for repeat victimization and gets away from the traditional method
of text-based database searches. A geographical reference greatly improves the chances
of an accurate ‘hit’ when searching for incidents of repeat victimizations, depending on
the accuracy of the geo-reference and how often the system is updated to cope with new
addresses. The 100-metre postcode grid references can coalesce, resulting in dozens of
addresses having the same postcode. Address-Point data are a considerable improve-
ment, but it is not always necessary to use the National Grid coordinates. Hirshfield and
others (1995) at Liverpool University have been analysing crime incident data using the
Merseyside Address Referencing System (MARS), which forms the basis of the
IDENTIFYING REPEAT VICTIMIZATION

Command and Control and Crime Incident reporting system used by Merseyside Police. This database was available to the researchers, as was that part of MARS which provided a one-metre resolution grid reference for each property in the study area. All premises are identified by Unique Property Reference Numbers (UPRN). The UPRN is recorded on every crime report and in their command and control data capture operation. Although the UPRN is unique, the grid references can be shared by different addresses. In a recent study (Johnson et al. 1997) the researchers had to write their own software to search for repeat victimizations. A FORTRAN program was constructed which combined a grid reference and address text-based search. Manual checking of the data for ambiguous results was a necessary final step. Although the MARS system allocates a one-metre resolution grid reference, a high resolution OS grid reference is not sufficient to deal with the address problem of multi-occupancy multi-storey buildings. A solution to this problem might be the Ordnance Survey Address-Point Reference (OSAPR) which identifies different addresses within the same building. The following study aims to demonstrate that standard GIS software is capable of quickly and accurately identifying repeat victimization events from Address-Point referenced crime data.

The Nottinghamshire Study

A commercial GIS and mapping package, MapInfo, was used for this study of Nottinghamshire Constabulary crime data. The data amounted to just over two years of burglary offences recorded on Trent Police Division, Nottinghamshire Constabulary. While burglary details are used in this study, police recorded crime data tend to contain more information than just the type of crime and the location. Nottinghamshire crime data include vehicle data for offences involving vehicles, burglary methods, points of entry and victim data such as ethnicity, age and gender. Using this extra information a more informative picture of repeat victimization might be created. This would enable the police to concentrate their crime prevention resources on ethnic minorities or the elderly if they wished. Commercial packages such as MapInfo or ArcView allow searches of the data to be made using a Standard Query Language input, and proprietary programming languages (e.g. MapBasic) allow many of the queries and functions described here to be automated.

Trent Police Division covers most of the South and East of Nottinghamshire and mixes affluent suburbs, council estates and rural villages within one police division. This study brought the focus more specifically to one of the station areas within Trent Division. The first search found every location where more than one burglary report had been recorded by selecting all records where there was at least one other record with an identical geographical location. Any other crime type could easily have been chosen. The nature of these locations tended to be domestic properties with a few small light industrial estates also in the study area. The Nottinghamshire CRIS aims to include the grid reference of the crime location, not the grid location of the victim (if they are different). Crimes that often happen in the street, such as assaults and personal robberies, are allocated the grid reference via Address-Point of the nearest building known to the system. Address-Point is aimed at identifying households. The use of this type of data means that the police have difficulty recording the location of any offences...
which occur outside these ‘known’ locations, such as in the street or on waste ground. The one-metre accuracy tends to place offences which occur in the street, in the kitchen of the nearest house!

At the time this study was completed (1997), Nottinghamshire Constabulary was in the process of changing from PAF to Address-Point data and some addresses in the county were still referenced in the crime data with PAF 100-metre resolution geo-references. We had to check the data found by the GIS to remove erroneous records where the PAF grid reference was the same, but the address referred to different properties. It was heartening to detect that out of the dozen errors detected, only two were related to errors in the Address-Point data, the remainder being due to the resolution of the more inaccurate PAF. Twenty-one repeat period records were missing from an expected 387, leaving 366 to form the basis of the analysis. These losses were due either to incomplete date entry recording of the original events, or errors in the data set.

A number of texts identify problems of integrating GIS and other technology into police forces (Hirshfield et al. 1995; Maltz et al. 1991; Openshaw et al. 1990). The use of graphical displays for viewing crime data makes the information more understandable and therefore accessible to often less technologically experienced police officers. The graphics make it immediately obvious where the majority of incidents are taking place on the sub-division. The example given here shows how such displays might be employed. Figure 1 shows details of the study area. The symbols represent the locations of all reported burglaries in the area from April 1995 to April 1997. The number of burglaries at each location is not shown, each dot representing an address that has been burgled at some point, irrespective of number of burglaries at that point. The data overlay a raster image of the street layout. There is a high incidence (Figure 1) of burglary across the study area, and the accurate targeting of crime prevention resources is certainly a local priority. The first step towards identifying repeat locations is to isolate the addresses which have been burgled more than once in the study period. This can be seen in Figure 2. It then becomes apparent that the distribution of repeat burglaries is different to the distribution of all burglaries in Figure 1. A large number of repeats appear concentrated around a busy road junction. It is also noticeable that in some areas of the map repeat burglaries are almost unknown. Possible reasons for this are that the residents took it upon themselves to improve their home security or other more complex social factors.

Figure 3 shows the same detail of the study area and the same locations where repeat burglaries have occurred. This map shows the magnitude of repeat burglaries at each address over the two-year period. The larger the circle, the more crimes have been reported to the police as occurring at the location. The largest circle (not shown in the figure) marks a council works depot which suffered seven burglaries over the two years, and demonstrates addressing problems discussed earlier. The depot is not currently included within the Nottinghamshire Constabulary address database and crimes at the location have been assigned to the nearest property, a dwelling house in the same street. This gives the inaccurate appearance of a high number of repeats at a domestic property.

Table 1 shows the distribution of victimization for all burglaries across the study area from April 1995 to April 1997. The data examined covered the whole of the time period. A minor correction for edge effects was therefore necessary. In some studies which have looked at burglary figures, the researchers have only been able to examine a number of months data (for example Anderson et al. 1995 who examined 11 months of data for their
IDENTIFYING REPEAT VICTIMIZATION

**Fig. 1** Detail of the study area. The symbols indicate locations which were burgled from the period April 1995 to April 1997. © Crown copyright. ED 273554

**Fig. 2** Locations where more than one burglary took place over the study period have been isolated. © Crown copyright. ED 273554
study). If burglaries repeat (for example) within one month, the events that happen at
the very beginning and very end of the study time period are at a statistical disadvantage.
Events at the beginning are denied the possibility of being repeats to crimes that
happened just before data recording commenced. Similarly crime events at the end are
denied possible repeats. The Nottinghamshire study looked at crime figures for a
two-year period, and the greatest repeat time between burglary events shown in the
graph (Figure 4) is a half-year period (26 weeks). In a two-year period (104 weeks), there
were only 78 weeks where a repeat burglary half a year later would be included in the data
available. If a repeat time of only two weeks is being recorded, there are obviously more

Table 1  Distribution of victimization for all burglaries in the study area: April 1995–April 1997

<table>
<thead>
<tr>
<th>Times burgled</th>
<th>Properties affected</th>
<th>Incidents</th>
<th>Incidents (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1356</td>
<td>1336</td>
<td>68.2</td>
</tr>
<tr>
<td>2</td>
<td>174</td>
<td>348</td>
<td>17.5</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>117</td>
<td>5.9</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>44</td>
<td>2.2</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>40</td>
<td>2.0</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>42</td>
<td>2.1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>14</td>
<td>0.7</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>16</td>
<td>0.8</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>0.5</td>
</tr>
</tbody>
</table>
weeks of data available which can contribute to the total. The correction used here is the same as employed by Anderson et al. (1995: 47), and prevents the time-course graph decreasing artificially (Equation 1.1). The effect of this correction can be seen in Figure 4.

\[
\frac{\alpha \eta}{\eta - \beta}
\]  

(1.1)

where \(\alpha\) is the number of events, \(\beta\) is the number of weeks repeat time, \(\eta\) is the total number of weeks in the survey (104).

There are problems with this formula related to the longer time frame used in this study which leads to over-compensation in the final months. The correction is more useful if restricted to shorter time scales.

There were 1,988 reported burglaries across the sub-divisional area, affecting 1,600 properties, during the two years. Over 68 per cent of these properties were never victims of burglary again during the study period. Just over 70 per cent (71.6 per cent) of the remaining properties were burgled one further time leaving 70 out of the 1,600 properties victimized more than twice. Whilst these 70 properties account for only 3.5 per cent of the number of victimized properties, they account for 14.3 per cent of the number of burglary events. As in other studies (Johnson et al. 1997), burglaries which did not have a clear event date and could have occurred over a number of days, such as a weekend, were allocated the average of the starting and ending date. When the repeat time course graph is examined (Figure 4) it can be seen that the time between burglaries tends to agree with other published work in that the greatest risk of a repeat is the time immediately after a burglary. However this study found that the magnitude of the initial repeat period was less severe than in other studies. Anderson and his colleagues (1995) found that 40 per cent of all repeats happen within a month of the preceding one, while Burquest et al. (1992) found an even greater figure of 79 per cent of revictimization incidents occurring within one month for school burglaries in Merseyside. These compare with a much less dramatic figure in this study of 27.8 per cent for the first month.
(up to and including 30 days) for repeats occurring within a year, or 23.3 per cent if repeats up to two years apart are included.

There could be a number of reasons for this difference. Anderson was looking at domestic burglaries exclusively while Burquest was looking only at school burglaries. This study did not differentiate between types of burglary and examined all burglaries, domestic or otherwise. When the GIS was employed to select locations which had been revictimized twice or more frequently it was found that over 80 per cent of the premises were non-residential. These included sports centres, schools and building sites. Police experience would suggest that thefts from building sites tend to have an extremely high reporting rate and that these crimes also tend to be reported as burglaries. Theft of tools from sites are reported by site foremen as the tools are often claimed back from the contractors or have to be accounted for. The loss of the tools is also preferentially reported as a burglary, as a theft might cast suspicion on work colleagues. Similarly schools and sports centres tend to have set guidelines for reporting burglaries and report a higher percentage of crimes that come to light.

The usefulness of geo-referenced data to identify repeat victimization is dependent on a number of factors. The problems already identified with the PAF would suggest that a database of crime data geo-referenced with only PAF coordinates could be used only as an initial search tool. A manual search of the address data would still be necessary to remove erroneously juxtaposed crime events. Address-Point data, or similar unique address references, would appear to be a substantial improvement on PAF and would suggest that the identification of repeat victimizations within police data can now be greatly enhanced.

The use of geo-referenced data would come into its own if the definition of repeat victimization was enlarged to include bigger geographical areas. Vehicle crimes such as theft from motor vehicle and theft of vehicles tend to occur in the street. Identifying a particular point on a street where a vehicle crime occurred, from police recorded crime data, is near impossible. The few studies that have identified repeat vehicle crimes have had difficulty in describing a repeat vehicle crime. Anderson et al. (1995) define three possibilities:

1. any vehicle, anywhere, from the same complainant;
2. any vehicle from the same location irrespective of owner;
3. the same vehicle from any location even across changes of owner (p.8).

Areas such as street corners where a number of roads meet and are poorly lit make parked vehicles an easy target. The crime reports to these events would show a number of different streets and would not be detected as a repeat by text-based search engines. The plotting of the geo-referenced crime data within a GIS would enable the operator to identify the street corner as a possible source of the trouble. An example of this is given in Figure 2, where many streets meet at a single junction in the delimited area. Although this is a case study of burglaries, motor vehicle crimes occurring at this street junction would be recorded as occurring at nearby properties and hence the power of a GIS is required to extract from the database all those vehicle incidents within some distance of the junction. This is one future direction for this type of work and gets into the territory of ‘hotspot’ analysis.

Standard GIS procedures have been used in this study to locate repeat victimization as an alternative to a macro written in a standard database to identify locations which shared
identical geo-referenced coordinates. The advantages of using a GIS over a standard database package are the ability to plot graphically the locations selected on a map for visualization of the problem, combined with the ability to select repeats around a particular geographical feature, such as a road junction or town centre.

If a local authority wished to invest crime prevention resources in a geographical area such as a run-down estate, a text-based system would have immense difficulty in identifying repeats in that area alone. Such areas tend to not fall into particular police boundaries and a non-geo-referenced search would either produce locations outside the desired search area, or worse, miss out locations within the estate. This is not a problem when a GIS is used as the area of the estate can be more precisely defined by the user. The GIS can then be tasked to detect repeats in the specific streets of the estate and only those streets. It can also select the correct streets without fear of selecting crimes from identically named streets in other areas of the city or town.

When is a repeat just another individual incident?

Many of the other studies referred to here cover shorter periods of between six months and a year. The longer study period of two years raises questions regarding the interpretation of repeat victimization. Repeats, by definition, have a relationship with the initial incident. This might be because the same burglar returns to the location, or he informs associates that the address is particularly vulnerable. At what point, therefore, do repeat incidents become unconnected with the initial occurrence? Is the burglary just another one which happens to be at an address where a previous incident once took place? Polvi et al. (1991) conducted a four-year study, one of the longest studies in the literature, but failed to address this point. The drop in risk decreased in the Polvi data from an initial high until six months after the initial incident, at which point the level of risk returned to the same level as for the rest of the study area. The graph presented by Polvi demonstrates the absence of any elevated risk after 6–7 months. Figure 4 shows a logarithmic curve fitted to the Nottinghamshire data. This curve crosses the x-axis before a year has passed, suggesting a similar decline in risk. Can repeats that occur a year or more after the initial incident truly be considered repeats, or are they new initial events?

In an attempt to explore this question the original data were re-examined. Three variables were identified which might give some indication that the repeat incident was committed by a criminal or criminals with a similar modus operandi to the original event. Nottinghamshire Constabulary record the Point of Entry (POE) where the burglar gained entry to the location (for example, upper first floor sliding window, or roof skylight), the Method of Entry (MOE) employed (for example, using a glass cutter, or bodily pressure), and the times between which the offence was committed. This enabled a crude index of crime similarity to be constructed. Repeat incidents were considered matched if they had an identical POE, if they had an identical MOE, or could be identified as both being daytime (or night time) burglaries. It soon became clear that errors or omissions in the data would seriously hamper the investigation. Table 2 shows the ratio of lost records.

The reporting process is a definite source of error in this type of data. The officer at the scene reports the burglary and returns to the station where he or she fills in a crime sheet. The crime sheets are then entered onto the database by a civilian operator, and not the
individual who visited the scene of the crime. The system therefore relies on the reporting officer accurately describing the burglar’s method and point of entry (if known). Absence of this information on the crime sheet means an empty relevant field. These were detected and removed from the data, as shown in Table 2. Nottinghamshire Constabulary have tried to remove this source of error by redesigning the crime recording sheets. However there is also the problem of interpretation of the burglary crime scene by the officer. With 169 different POE options to choose from, it is likely that different officers will interpret and record the same POE differently from time to time.

Initial analysis showed that from the 236 records in the data, 113 exhibited a match between occurrence and repeat in terms of either time of day, MOE or POE. It is assumed in this section that where a match exists, there is a reasonable probability that the repeat was directly related to the previous incident. If no match was found, as in 123 of the cases, this could suggest that the repeat was independent of the occurrence, and the apparent relationship between incident and repeat may not exist. A possible hypothesis is that the longer the repeat interval the higher the probability that a repeat is not a genuine repeat but a fresh incident.

There may be a number of reasons for a repeat not being a repeat beyond the criminal being a different individual. The environmental factors which highlighted the property as vulnerable may have changed such as improved crime prevention measures or different policing patterns. This environmental change might be reflected in a different *modus operandi*. An attempt to investigate this possibility is shown in Table 3, where the ratio of matched repeats to total repeats for 30-day periods was calculated up to a six-month limit. In this data set the number of repetition burglaries becomes too small to be reliable beyond six months. Over the six-month period, however, there is a fairly steady decline in the proportion of matched repeats. This suggests that as lag-time increases the number of genuine repeats detected within the data declines.

<table>
<thead>
<tr>
<th>Months between repeats</th>
<th>Ratio of similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>0.61</td>
</tr>
<tr>
<td>3</td>
<td>0.59</td>
</tr>
<tr>
<td>4</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>0.56</td>
</tr>
<tr>
<td>6</td>
<td>0.45</td>
</tr>
</tbody>
</table>
Conclusions

This study has shown the benefits of using GIS to detect repeat victimization. The accuracy and speed of analysis were greatly enhanced by using police crime data which included full Ordnance Survey geocoding directly from address data. This led to great improvements in both overall processing time and final accuracy of the detected repeats. The repeats formed less than a third of the total crime data set for the study station area on Trent Division, Nottinghamshire Constabulary. From this limited data set it would appear that, using the additional data commonly found in police records, it is possible to make a first approximation as to the temporal extent of genuine repeat burglaries. Research is underway to determine whether this method is applicable to larger data sets and hence whether it has implications for proactive crime prevention procedures. Policing policy will differ for repeat victimization and multiple victimization. In the former case with similar MOs specific measures can be implemented. In the latter, where a variety of techniques and criminals are found, a more general policy may be necessary.

References

Mortem of a failed GIS’, 2nd National Conference and Exhibition of the AGI, Brighton.