Statistical modelling in specific case analysis

CGG AITKEN

Department of Mathematics and Statistics, The University of Edinburgh, Edinburgh, United Kingdom EH9 3JZ

T CONNOLLY

Biomathematics and Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee, United Kingdom DD2 5DA

A GAMMERMAN

Department of Computer Science, Royal Holloway and Bedford New College, University of London, Egham, Surrey, United Kingdom TW20 0EX

G ZHANG

Department of Electrical Engineering and Computer Science, Lehigh University, Bethlehem, Pennsylvania 18015, United States of America

D BAILEY, R GORDON

Derbyshire Constabulary Headquarters, Butterley Hall, Ripley, Derbyshire, United Kingdom DE5 3RS

and

R OLDFIELD

Police Research Group, Home Office, Queen Anne's Gate, London, United Kingdom, SW1H 9AT

Science & Justice 1996; 36: 245–255

Received 31 July 1995; accepted 9 January 1996

In spite of the problems associated with the development of a statistical approach in specific case analysis, the potential of two statistical methods, logistic regression and Bayesian belief networks, has been investigated and found encouraging in the context of a database relating to child murders with a sexual connotation. Continual collaboration between statisticians and detectives would be needed in the collection of the data, the choice and construction of the models, and the interpretation of the results.

Die Entwicklung von statistischen Verfahren zur Betrachtung von Täterprofilen ist bekanntlich mit Problemen verbunden. Trotzdem wurde an einer Datenbank über Kindsmorde mit sexuellem Hintergrund die Anwendbarkeit von zwei statistischen Mehtoden -dem „Logistic regression model" und dem „Bayesian belief network"- untersucht. Die Erbenisse waren ermutigend. Künftig wäre aber eine kontinuierliche Zusammenarbeit von Statistikern und Kriminalbeamten bei der Erfassung der Daten sowie der Auswahl und Entwicklung der statistischen Modelle und der Interpretation der Ergebnisse notwendig.

Malgré les problèmes associés au développement d'une approche statistique dans le profil de délinquants, le potentiel de deux méthodes statistiques, la régression logistique et les réseaux de croyance Bayésien ont été investigés dans le contexte d'une base de données de meurtre avec connotation sexuelle d'enfants. Les résultats sont encourageants. Une collaboration continue entre les statisticiens et des détectives serait nécessaire dans la collection des données, le choix et la construction des modèles et l'interprétation des résultats.

A pesar de los problemas asociados con el desarrollo de un acercamiento estadístico al perfil del delincuente, el potencial de los dos métodos estadísticos, la regresión logistica y el método Bayesiano, se ha investigado en el contexto de una base de datos referente a asesinatos de niños con connotación sexual, y los resultados fueron alentadores. La continua colaboración entre estadísticos y detectives seria necesaria en la recogida de datos, la elección y construcción de modelos, y la interpretación de los resultados.

Key Words: Forensic science; Bayesian belief networks; Logistic regression; Model assessment; Specific case analysis; Statistical modelling.

Introduction

Specific case analysis or 'offender profiling' is the name given to a collection of scientific and psychological theories and techniques that seek to relate offender, victim and crime characteristics, in order to assist the police with crime investigation.

Traditionally, detectives have had to rely on their own common sense, intuition, expertise and experience to investigate serious crimes. More recently, psychologists have applied the theories which have been developed in universities or hospitals to live crime investigations. Work in this area of behavioural sciences has led to some remarkable success in helping police with their enquiries [1,2], although it was not thought feasible to use statistical techniques because of the lack of any significant volume of suitably detailed datasets.

However, in 1986 work was begun on the construction of a dataset containing details of all sexually-motivated child murders and abductions which had been reported in Great Britain since 1960. This work has continued and by 1991 the dataset held details of characteristics of the offender, where known, of the victim, and of the crime scene in over 400 cases. The possible use of such a dataset to provide even a basic probabilistic description of an offender was of considerable interest. The approach to be discussed here has analogies with epidemiological studies and, in particular, with so-called case-control studies as used in the human and veterinary medical fields. Factors which might be thought to be associated with a disease in humans or animals are investigated in two groups of humans or animals, one group with the disease and the other group without it. Other applications [3–5] of epidemiological methods to criminology have been concerned with seeking associations amongst variables. The aim of this paper is to investigate methods for making probabilistic statements about characteristics of the offender, based on observations from the crime scene and the victim. The present work, however, was not concerned with the evaluation of evidence in support of the guilt or otherwise of a suspect, but with a different aspect of a criminal investigation, that of helping police with their enquiries and determining priority paths for their investigation.

It is the purpose of this paper to show that there is a role for statistics in the prediction of offender characteristics using the known characteristics of the victim and information from the scene of the crime, and to discuss problems associated with this role. The methods reported in the paper are an attempt to establish a statistical approach to offender profiling. The methodology was developed with respect to a particular class of crime, namely child murders with a sexual connotation, hereafter known as relevant murders, in which there was a single offender and a single victim. The methods described will be applicable to other classes of crime in which there is a single offender and a single victim, but extensions to serial crimes have not been investigated.

Statistics and offender profiling

Relationships between offender characteristics and victim and crime characteristics can be investigated using statistical techniques. In particular, there are procedures which may be used to provide a probabilistic prediction of the characteristics of an offender from the victim and crime characteristics of individual cases. The predictions are presented in such a manner that the probabilities are associated with the values of a particular characteristic. The performances of the different methods may be assessed, primarily through the calculation of the proportion of predictions which are wrong, i.e. the misclassification rate. A low misclassification rate is associated with a good method.

The techniques to be discussed here are independent of any particular dataset being studied, provided that the dataset fits within a very general framework. This framework is the same whether one is considering the relevant murders of this paper or rape data or many other types of crimes. There may, though, be considerable difference of detail between different types of crime. For example, in many rape cases it will be possible to take a statement from the victim, something which is obviously not possible in murder cases. The framework is such that there are certain characteristics, the values of which are to be predicted, and there are others which may be used to make the prediction. In the context of offender profiling, the characteristics of the victim and crime scene are used to predict those of the offender.

One problem is the assessment of the quality of the discrimination of a particular offender from the total offender population, based on certain crimes, or from the general population. Information on the latter may be gained by referring to census figures concerning age, sex and marital status, for example, but for this to be most useful, a complete match between categories in the database and census categories is required. Consideration also needs to be given as to what group should be taken as a control group for a certain crime. Data are required from the control group in the same form as is used for the crime under investigation (e.g., child murders, rapes).

Once a control group has been identified, analysis may be performed using a case-control study. The cases are the offences for the crime under investigation, with the details on the database. The controls are a group of people, similar to the offenders on the case database, with whose characteristics the offenders can be compared. Differences in characterisation may then be inferred to arise from the distinction that people in the case database are offenders in the relevant crime, whereas people in the control database are not (though they may be offenders in other types of crime). These differences may then help in future crimes to identify particular groups of people to which the offender may belong. The classification of observations into cases and controls can be done in several different ways and needs to be a matter for discussion amongst all interested parties.

First, cases can be offenders for the crime of interest and controls can be a general population for which information can be gathered from census records. Alternatively, a choice may be made of a more specific population from which the offender may be thought to have come. For example, it may be that in crimes of rape, the control population is not those of innocent people but is those of burglars [6]. Burglars would be chosen because rape and burglary are similar with respect to seriousness, the violation of personal space and aggressiveness, often towards women. The difference between the groups is in the act of rape itself. Any differences in the responses to the various characteristics considered can then be ascribed to the fact that one group raped and the other did not. In general, given the circumstances of a crime, probabilistic predictions of the characteristics of the offender can be made from the case group and these predictions may be compared with the distribution of the same characteristics within the control group. It is only then that the usefulness of the probability statement can be properly assessed. Suppose a probabilistic prediction were desired of the marital status of the offender in a particular crime. Consideration of the victim and crime scene characteristics predict that, with probability of, for example, 0.75, the offender is living with a partner. The odds of 3 to 1 on that the offender is living with a partner may be thought useful, but this result needs to be compared with the proportion of people in a general population who live with a partner. Suppose this probability is also 0.75, the odds are also 3 to 1 on that the person, chosen at random from this population, is living with a partner. In this instance, consideration of the crime and victim characteristics has led to no new information. (Note, however, that these figures are purely hypothetical; in the database for the crimes considered here, only 24% of offenders were living with a partner.)

Secondly, cases can be offenders with a particular characteristic and controls can be offenders without this characteristic. Thus, as described later, in predicting the relationship between the offender and the victim in a child murder, cases could be offenders who were known to the victim and controls could be offenders who were unknown to the victim.

Unfortunately, control data were unavailable in this study and the two models discussed in the following sections were based solely on case data.

The data

The database used in this paper was one compiled for all child sexual murders in Great Britain from 1960. Full records were available from 1960 to 1991 inclusive; results in the present work relate to this subset of records. The data, which exist in coded form, were obtained by constructing a pro forma listing all details which might be thought relevant. Recorded details related to characteristics of the offenders and of the victims and crime scenes. The choice of which characteristics to include was made by the detectives collating the data. Each case record was then studied and the details were transferred to the pro forma, entered into a computer database and stored.

The data under consideration were restricted to those cases in which there was a single offender and a single victim. Undetected cases, serial offenders, and cases with multiple offenders and/or victims were excluded. A study [4] of so-called primary homicides (ones which did not involve the perpetration of another crime) only included the first specified offender and the first specified victim in the analysis and thus ignored evidence from other crimes in the series. Concentration on cases in which there was a single offender and a single victim left 320 cases for analysis. The excluded cases may involve different motivations and different analyses may be required. Further, the use of cases with only a single offender and a single victim ensured independence of the cases, which was of importance when carrying out the statistical analyses. Independence may not hold when serial offences are considered, for example.

Offender characteristics

When the construction of a new dataset is being considered, there is a need for an initial discussion about the choice of offender characteristics. For the relevant murders studied here, the sex of the offender was not of interest since, with only two exceptions, the offender was male. For the two exceptions, the female offenders were accomplices of male offenders.

Age is a continuous variable though in practice only a crude classification is of interest. For example, the offender may be given a binary classification of young (0–20) or old (21 plus) as in the study described here.

The marital status of the offender was originally recorded as four categories: married, single, single and divorced, single and cohabiting. However, for the purposes of this research, only two categories were considered: living with a partner (married or single and cohabiting) or not living with a partner (single or single and divorced).

The offender's relationship with the victim was recorded as one of five categories: parent, cohabitee, relative, acquaintance and stranger, ordered according to a perception of familiarity. The analysis reported here used a binary classification and the offender was classified as either known to the victim (parent,cohabitee, relative or acquaintance) or not known (a stranger).

The proximity of the point of contact of the offender and the victim to the normal residence of the offender was classified as a binary variable: within or outwith a five-mile radius of the point of contact. In fact, over 90% of offenders lived within five miles of the point of contact so prediction of this variable did not provide much additional information.

For those offenders who had previous convictions, these were recorded in 11 defined categories and one labelled ‘Other’. The defined categories had been thought to be particularly relevant by the compilers of the data to the crime in question, namely murders of children with sexual connotations. However, it was discovered through a simple count that most offenders with previous convictions had them in the ‘Other’ category. It was not possible from the data recorded to be more explicit. It is of considerable interest that a previous history of crimes such as rape, unlawful sexual intercourse and incest were not particularly common amongst child murderers. The analysis here used the binary classification: preconvictions, yes or no.

For the purposes of the study, these five offender characteristics were considered, separately, with two binary levels for each.

Victim/crime scene characteristics

For each offence, over two hundred characteristics were recorded about the victim and the scene of the crime, which could conveniently be arranged into groups which were: (1) victim age and sex, (2) victim location at point of contact, (3) offender activity, (4) crime locations, (5) activities during offence, (6) distances and proximities, (7) times, eg, day of week, time of day, and (8) offender peculiarities. It is not possible to give details of the characteristics which were used in the final analysis but certain points of general interest can be made. Three victim variables, sex, age and abduction (whether the victim was abducted or not) were of particular interest as these were notable features of the cases which led to the formation of the database.

The cross-tabulation of age and sex of the victims (Table 1) showed a highly significant association between age and sex. The chi-squared test of no association gave $X^2 = 11.4$ on 1 d.f., $P < 0.001$. The odds ratio was 2.3, and the interpretation of this association was that for a male victim the odds in favour of being under 11 years old (46/43) were more than twice those of a female victim (102/218).

The cross-tabulation of age and abduction (Table 2) also showed a highly significant association ($X^2 = 16.5$ on 2 d.f, $P < 0.001$). For victims under 11 years old, 25% were abducted, compared to 10% of the victims aged 11 years and older. No statistically significant association was found between the sex of the victim and whether the victim was abducted or not. If the 56 undetected cases in Table 2 were ignored, there was an odds ratio of 3.0 in the remaining table; i.e., amongst detected offences, for a victim under 11 years old the odds in favour of abduction were three times as great as for a victim of 11 years or older.

Statistical models

Logistic regression

An appropriate statistical model for the prediction of the level of an offender characteristic in this context is one known as the logistic regression model. Here the log odds in favour of the offender being in one of the levels of the characteristic is expressed as a linear combination of certain victim and crime scene characteristics, and gives an equation known as a logistic regression equation. The offender characteristics considered to be of interest were taken to be binary and these, and the categories (levels) into which they fell, are shown in Table 3.

Let p be the probability of the offender taking the first level of a particular offender characteristic; then $1\text{-}p$ is the probability of the offender taking the second (other) level of this characteristic. Let there be k victim and crime scene characteristics $x_1,\ldots,x_k$ in the model. These are also taken to be binary at present. Thus x_j ($j=1,\ldots,k$) takes one of two values, 0 and 1, corresponding to the two levels of the characteristic. For example, consider the age and sex of the victim. The binary record of victim age has ‘age 0–10’ as the first level and ‘age 11 or over’ as the second level. The binary record of sex has ‘male’ as the first level and ‘female’ as the second level. If age and sex were the only two characteristics considered then $k = 2$ and let x_1 correspond to age and x_2 correspond to sex. Then $x_1 = 0$ would correspond to a victim ‘age 0 –10’ and $x_1 = 1$ to a victim ‘age 11 or over’; $x_2 = 0$ would correspond to a male victim, $x_2 = 1$ to a female

TABLE 1 Cross-tabulation of victim sex with victim age.

Victim Sex	*Age (yrs)* *0–10*	*11+*	*Total*
Male	46	43	89
Female	102	218	320
Total	148	261	409

TABLE 2 Cross-tabulation of victim age with victim abduction.

Abduction	*Age (yrs)* *0–10*	*11+*	*Total*
Yes	37	26	63
No	92	198	290
Undetected	19	37	56
Total	148	261	409

TABLE 3 Offender characteristics, levels and the level in favour of which the log odds are derived.

Variable	*Level* *First*	*Second*	*Log odds in favour of:*
Age	0–20	>20	0–20
Marital status	Living with partner (Lwp)	Not living with partner	Lwp
Relationship to victim	Known (Parent, relative, cohabitee, acquaintance)	Stranger	Known
Proximity of crime to offender's normal place of residence	Within 5 mile radius (W5mr)	Outwith 5 mile radius	W5mr
Previous convictions	Yes	No	Yes

victim. The logistic regression equation links the offender and victim characteristics thus:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \sum_{j=1}^{k} \beta_j x_j \qquad (1)$$

where α is a constant and $\beta_1, \ldots, \beta_k$ are the coefficients, known as regression coefficients, with which to multiply the observed values (0 or 1) of $x_1, \ldots, x_k$, respectively. Thus, if x_j is at the second level ($x_j = 1$), $\log\{p/(1-p)\}$ is adjusted by the addition of β_j; if x_j is at the first level ($x_j = 0$) $\log\{p/(1-p)\}$ is unaltered. The right-hand-side of (l) may be thought of as a score. A value of $x_j = 1$ adds β_j to the score. In (l) the left-hand side is the 'log-odds' and the right-hand side is the linear combination mentioned above.

In this example, if a boy aged 8, say, and thus in the age range 0–10, was murdered, $x_1 = 0$, $x_2 = 0$ and $\log\{p/(1-p)\} = \alpha$. If a girl aged 12 (and thus with an age '11 or over') was murdered, $\log\{p/(1-p)\} = \alpha + \beta_1 + \beta_2$. From these equations, p may be determined. For example, from the first case,

$$p = \frac{exp(\alpha)}{1 + exp(\alpha)}.$$

A stepwise method in the SAS LOGISTIC procedure [7] was used to choose the victim and crime scene characteristics, including the number, k, of characteristics and the corresponding values for α and $\beta_1, \ldots, \beta_k$, to be included in the final model for each offender characteristic. This choice was made from those victim and crime scene characteristics chosen by the user (here, the compiler of the database) initially. A comparison was made of models fitted with different numbers of characteristics. The final choice of the model to use was that one which provided the best fit consistent with containing a reasonably small number of variables. There were 118 characteristics selected for possible inclusion in the models from the original 200. Characteristics excluded were those that were present in less than 5% of the cases under investigation.

A separate logistic regression model was determined for each offender characteristic. The victim and crime scene characteristics chosen for each model were those which discriminated best between the two levels of the offender characteristic.

A two-by-two table, in which a decision has to be made between the first level A and the second level $\bar{A}$ of a characteristic, may be useful in understanding the definitions of the measures of performance given below. In Table 4, there are n cases in total, of which $a + c$ are of type A and $b + d$ are of type $\bar{A}$. However, it is decided using a statistical model that $a + b$ are of type A, for a of which this is the correct decision and for b of which it is the wrong decision. Similarly it is decided that $c + d$ are of type $\bar{A}$, for c of which this is the wrong decision and for d of which it is the correct decision.

The measures of performance are defined as follows, with the corresponding relative frequencies from Table 4 given in parentheses:

Correct: the percentage of the total number of cases that were correctly predicted $\{(a+d)/n\}$;
Sensitivity: the percentage of the cases with the first level of the response variable that were correctly predicted $\{a/(a+c)\}$;
Specificity: the percentage of the cases with the second level of the response variable that were correctly predicted $\{d/(b+d)\}$;

TABLE 4 A general two-by-two table for definitions of statistics used in measuring performance.

Decision	*Truth* *A*	$\bar{A}$	*Total*
A	*a*	*b*	*a+b*
$\bar{A}$	*c*	*d*	*c+d*
Total	*a+c*	*b+d*	*n*

False positive: the percentage of the cases predicted at the first level of the response variable which have been incorrectly predicted as such $\{b/(a+b)\}$;
False negative: the percentage of the cases predicted at the second level of the response variable which have been incorrectly predicted as such $\{c/(c+d)\}$.

Measures of performance for the five models are given in Table 5. The results show that the logistic regression model performed well with these data.

Example 1
During the course of the research, a crime of a similar type to the cases under investigation was committed and for which a person was found guilty. The details were entered into the logistic regression models and it can be seen from Table 6 that the models provided very good predictions.

Bayesian belief networks
A Bayesian belief network is a graphical representation of the relationships amongst the various characteristics, offender, victim and crime. In this context, a graph is a set of nodes and directed arcs. Each node represents a particular characteristic; two nodes are linked by an arc whose direction represents a causal or influential relationship. The absence of an arc between two nodes implies that the two characteristics associated with these nodes are conditionally independent of each other, that is, they are independent conditional on knowledge of the values of the other characteristics. There is also a restriction that the directed arcs cannot form a closed loop, so that it cannot be possible to start from a particular node and follow arcs to return to that node. For previous relevant discussions of Bayesian belief networks see [8] and [9].

Suppose characteristic A causes or influences characteristic B (Figure 1). For example, characteristic A may be 'Offender is living or is not living with a partner'. Let a denote 'offender living with a partner' and $\bar{a}$ denote 'offender is not living with a partner'. Characteristic B may be 'Sex of victim'. Let b denote 'victim is male' and $\bar{b}$ denote 'victim is female'. The strength of the influence is represented by two conditional probabilities $Pr(b \mid a)$ and $Pr(b \mid \bar{a})$, namely the probability that b (the victim is male) is true if a is true (the offender is living with a partner) and the probability that b (the victim is male) is true if $\bar{a}$ is true (the offender is not living with a partner). Notice the use of the vertical bar $\mid$ as a piece of notation. What is given to the left of the vertical bar is the characteristic about which a probability statement is desired. What is given to the right of the vertical bar is the information which is known. Thus $Pr(b \mid a)$ is read as 'the probability that b is true (the victim is male) given a is true (the offender is living with a partner)'. Estimates of these probabilities may be obtained from the database. The complementary probabilities are the probability the victim is female if the offender is living with a partner ($Pr(\bar{b} \mid a)$) and the probability the victim is female if the offender is not living with a partner ($Pr(\bar{b} \mid \bar{a})$). These latter probabilities do not need to be obtained from the database as they may be obtained by subtraction from 1 of the appropriate one of the first two probabilities given above. The particular implication of this two-node network is that the sex of the victim is affected by the marital status of the offender. An offender who is living with a partner may have a different sexual preference for his victim from one who is not living with a partner.

TABLE 6 Predicted and actual outcomes of the five offender variables of interest, for Example 1.

Predicted outcome	*Probability*	*Actual outcome*	*Correctness*
Age 0–20	0.65	21	Reasonable
Single	0.91	Single	Correct
Known to victim	0.96	Acquaintance	Correct
Within 5 miles	0.79	Yes	Correct
Preconviction	0.92	Yes	Correct

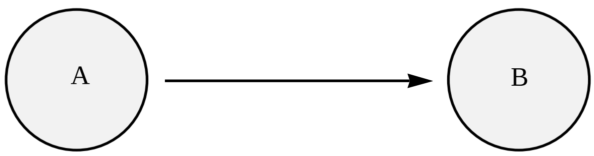

FIGURE 1 Diagrammatic representation that attribute *A* causes or influences attribute *B*.

TABLE 5 Measures of performance (%) for the five offender binary characteristics.

Characteristic		*Measure of performance (%)*			
	Correct	*Sensitivity*	*Specificity*	*False +*	*False -*
Age (0–20)	71.6	72.3	71.0	34.9	22.6
Marital status (Lwp)	75.9	21.4	95.3	37.9	22.7
Related (Known)	72.5	78.9	63.7	25.1	31.2
Proximity (W5mr)	90.3	96.6	22.2	6.9	62.5
Preconviction (Yes)	74.7	98.3	9.4	25.0	33.3

In practice, the reverse implication is required. Namely, it is of interest to know, given the sex of the victim, the probability of the offender living with a partner. It is a highly attractive property of the system that a suitable mathematical theory has been developed so that such implications can be made from the network.

A little terminology is appropriate here. The node from which the arrow departs, i.e., at the tail of the arrow is called a *parent* node. The node to which the arrow is pointing, i.e., at the head of the arrow is called a *child* node. Thus, it is said that characteristic A is the parent of characteristic B and B is the child of A. The methodology is best described by means of an example.

Seven node network

A network consisting of seven nodes was constructed in three stages in discussion with a senior detective. The first stage involved the specification of the characteristics to be included, i.e., the choice of nodes. The network was designed to be used to predict the offender characteristics (values for nodes 5, 6 and 7) from the victim and crime scene characteristics given (values for nodes 1, 2, 3 and 4). The description of the seven nodes and their possible levels are shown in Table 7. It should be noted that the restriction to binary characteristics has been removed so certain characteristics may have more than two levels.

The second stage constructed the relationships amongst the characteristics, i.e., which characteristics (nodes) were to be joined by arcs, as shown in Figure 2. Once this network had been constructed, the third stage, the specification of the conditional probabilities implied by the relationships, i.e., which conditional probabilities were required, became apparent.

For example, *victim sex* (characteristic B of Figure 1) was a child of *offender marital status* (characteristic A of Figure 1). Thus $Pr(b \mid a)$ and $Pr(b \mid \bar{a})$ would be required. These probabilities might be obtained from the database or, importantly, might be representatives of measures of the beliefs of an expert or group of experts. This is a particularly attractive feature of the approach. It is able to accept probabilities determined either objectively from a database or subjectively from the opinions of experts. It was also possible from the network to see what characteristics were parents or children of which other characteristics. For example, nodes 2, 3 and 4 were characteristics which had no children; nodes 5, 6 and 7 had no parents; and node 1 was the only characteristic which was both a parent and a child.

TABLE 7 Description of nodes and possible levels.

Node	*Description*	*Level*
1.	age of victim	0–7, 8–12, 13+
2.	sex of victim	male, female
3.	location of last sighting	home, other
4.	method of killing	strangulation, other
5.	marital status of offender	living with partner, other
6.	relationship of offender to victim	known, stranger
7.	preconviction status of offender	yes, no

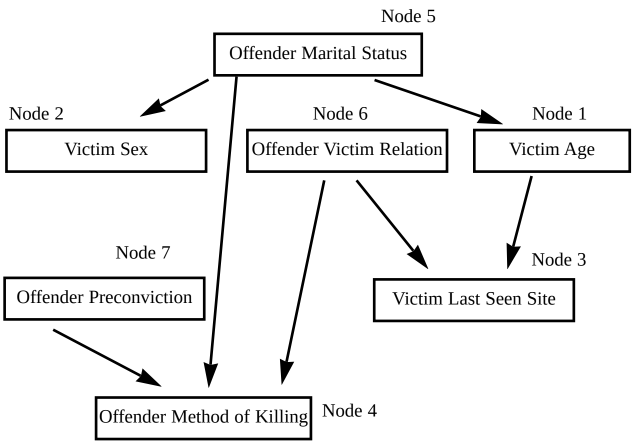

FIGURE 2 Seven node network showing the relationships amongst three offender characteristics and four victim and crime scene characteristics.

Probabilities were then determined. For nodes with no parents (nodes 5, 6 and 7), all of which were binary, one probability was required for each, the complementary probabilities being determined automatically once the original probability was known. The database gave 24% of offenders who were living with a partner and 76% who were not. Thus, for node 5, the probability that an offender was living with a partner was taken as 0.24, and the probability that he was not living with a partner was taken as 0.76.

For nodes with parents, conditional probabilities were determined. For binary nodes, one probability was required for each possible set of combinations of levels of characteristics amongst the parent nodes. Again, the complementary probabilities were determined automatically. However, node 1 was categorised at three levels. Thus, two probabilities were required for node 1 (the third follows automatically), for each of the two levels of the offender's marital status (node 5).

As an example, consider the sex of the victim. According to the network, this was dependent on the marital status of the offender. From the database, it was determined that the probability the victim was male, given the offender was living with a partner was 0.24, and the probability the victim was male, given the offender was not living with a partner was 0.13. The corresponding complementary probabilities were the probability the victim was female, given the offender was living with a partner was 0.76, the probability the victim was female, given the offender was not living with a partner was 0.87.

These probabilities implied that if the offender was living with a partner, the odds were 76/24 $\simeq$ 3 to 1 in favour of a female victim and if the offender was not living with a partner, the odds were approximately 7 to 1 in favour of a female victim.

Notice, however, that these probabilities are not what is required in the course of a police investigation. What is required is the transpose of these, namely if the victim is male (or female) what can be said about the marital status of the offender? The computer system enables such probabilities to be determined automatically. The system is a generalisation of Bayes' Theorem for conditional probabilities to consider the probabilities implicit in the network and is based on techniques described by Lauritzen and Speigelhalter [10].

The procedure used by the system may be illustrated using the example of the prediction of the offender's marital status from the sex of the victim. The probability of interest is the probability the offender is living with a partner, say, when it is known that the victim is male. Remember that *A* is the offender's marital status and *B* is the sex of the victim, with *a* denoting 'offender living with a partner', $\overline{a}$ denoting 'offender not living with a partner', *b* denoting 'victim is male' and $\overline{b}$ denoting 'victim is female'. Bayes' Theorem enables the following statement relating probabilities to be made.

$$Pr(a \mid b) = \frac{Pr(b \mid a) \times Pr(a)}{Pr(b)} .$$

From the figures given above, $Pr(b \mid a) = 0.24$, $Pr(a) = 0.24$, $Pr(b \mid \overline{a}) = 0.13$, $Pr(\overline{a}) = 0.76$. Also,

$$Pr(b) = Pr(b \mid a) \times Pr(a) + Pr(b \mid \overline{a}) \times Pr(\overline{a}) .$$

Hence,

$$Pr(b) = 0.24 \times 0.24 + 0.13 \times 0.76 = 0.1564.$$

Thus,

$$Pr(a \mid b) = (0.24 \times 0.24) / 0.1564 = 0.37;$$

i.e., if the victim is male, the probability the offender is living with a partner is 0.37. (Compare this with the probability that, if the offender is living with a partner, the victim is male is 0.24.)

Similar but more sophisticated calculations can be done to determine probabilities when there are more than two nodes involved in the conditioning. In practice, as information comes in to an investigation, the network can be updated by updating the probabilities of those attributes about which no information is currently available.

Examples

Suppose a crime has been committed and information about the victim and the crime scene is provided. The system, based on procedures described by Lauritzen and Speigelhalter [10] provides for a rapid propagation of this information through the network and the production of a revised set of probabilities for the offender characteristics. Two examples are given here.

Example 2

A female victim, aged 0–7 years, has been found strangled in her own home. The initial probabilities for the offender characteristics before receipt of this information and the probabilities revised after receipt of this information, as determined from the network, are given in Table 8. Notice that the probability the offender is living with a partner has increased slightly from 0.24 to 0.33, and the probability the offender is known to the victim has increased from 0.57 to 0.66.

Example 3

A female victim, aged 0–7 years, has been found strangled outwith her own home. The initial and revised probabilities for the offender characteristics as determined from the network are given in Table 9. Notice that the probability the

TABLE 8 Probabilities for offender characteristics for a female victim, aged 0–7 years, found strangled in her own home.

Characteristic	*Outcome*	*Probability* *Initial*	*Revised*
Living with partner	Yes	0.24	0.33
	No	0.76	0.67
Relationship	Known	0.57	0.66
	Unknown	0.43	0.34
Preconviction	Yes	0.73	0.73
	No	0.27	0.27

TABLE 9 Probabilities for offender characteristics for a female victim, aged 0–7 years, found strangled outwith her own home.

Characteristic	*Outcome*	*Probability* *Initial*	*Revised*
Living with partner	Yes	0.24	0.36
	No	0.76	0.64
Relationship	Known	0.57	0.11
	Unknown	0.43	0.89
Preconviction	Yes	0.73	0.70
	No	0.27	0.30

offender is living with a partner has increased from 0.24 to 0.36, and the probability the offender is known to the victim has decreased considerably from 0.57 to 0.11. The probability the offender has preconvictions has decreased very slightly. The only difference in the victim and crime circumstances from Example 2 is that the victim was found outwith her own home. This appears to have had a considerable effect on the probability of the offender being known to the victim.

Further technical discussion of Bayesian belief networks and an example of a ten node network are given in [11].

Discussion

The outcome of the application of statistical techniques to the particular database discussed here, while of considerable interest in its own right, illustrates the relevance of such techniques to offender profiling in general. One of the consequences of the statistical investigation was the suggestion of changes to the choice of characteristics and categories originally made by the detectives collating the data. However, despite these suggestions, very good results were achieved and the original data may be considered to be of high quality.

Lack of control data for this study meant that probabilistic predictions have therefore to be tempered with the caveats mentioned above about control groups. The results, though, were sufficiently encouraging for further work to be pursued. Included in this further work should be consideration of control groups.

The work on the database has raised a number of issues. Choice of categories was sometimes limited. For example, for those offenders who had had previous convictions, details of convictions which are currently classified as 'Other' could be noted. It was found that certain of the categories of the victim characteristics were ambiguous. There was a risk that the same characteristic could be counted twice. For example, two of the characteristics were the 'specific location of a victim at the time of last sighting' and 'the specific location of the contact with the offender'. Among categories recorded for these were 'at home or vicinity' and 'on a public thoroughfare'. A victim last seen on the pavement outside her home could be said both to be in the vicinity of home and on a public thoroughfare.

Some categories need to be more refined. In the study, for example, 46% of the cases involved the victim 'on a journey'. A finer definition of a journey is required. Certain characteristics were negative in nature or unclear. For example, 'inducement not known' was considered as a characteristic of interest in some of the models provided by the selection process. Another example was the use of a 'non-specific article'. Clear guidelines need to be given regarding interpretations.

These last two points regarding ambiguity and the negative nature of the characteristics would become very important if the pro forma were to be made widely available. The pro forma was completed by one individual, the compiler of the database, and the categorisation was consistent. The high quality of the categorisation of this is illustrated by the high quality of the measures of performance (Table 5).

Other issues include the necessity for internal consistency and for reproducibility. Internal consistency requires that the information on the database should be consistent both within and between characteristics. Reproducibility requires that a parallel analysis would give rise to broadly similar results and inferences. This is something which could be borne in mind in future investigations.

The five binary offender characteristics were considered separately, because there were insufficient data to do more. For example, if age, sex and abduction of the victim were considered there would be five binary offender variables and three binary victim variables with 2^8 (= 256) possible combinations and only 320 cases from which to choose the best model. Of course, certain of the offender characteristics may be related and future work could investigate this more thoroughly than the current authors were able to do, given the resources available.

The choice of 118 characteristics out of the original 200 for possible inclusion in the models was a subjective decision by the authors and the effect of varying this figure could also be the subject of further work. The intention was to exclude rare events which might have an undue influence on the model. For example, in only one case was a victim picked up at, or in the vicinity of, a known vice area. The number of cases for consideration was 320. The final models chosen for the five offender characteristics had numbers of included victim and crime-scene characteristics ranging from 5 to 13, a considerable reduction from the original 118. The final choice of characteristics to be included in a model was an entirely objective procedure, after the initial subjective selection of the 118 characteristics. No account was taken of detective expertise. This led to apparent anomalies in that the sex of the victim was only included in the model concerning the offender's marital status. In future work it is hoped that certain characteristics will be included as a matter of course, after consultation with detectives. Such a procedure would allow for the inclusion of subjectivity; different detectives could request the inclusion of different variables.

The logistic regression technique will carry over immediately to any problem in which there is an individual binary characteristic (for example, previous convictions: yes or no) to be predicted. The characteristics considered from the victim and crime scene do not need to be binary though the analyses are easiest when they are so. The cases under consideration may require to be mutually independent. As an example of dependent cases, consider serial crimes where data relating to crimes committed by the same offender on different victims would be thought to exhibit associations which would not be present in data relating to crimes committed by different offenders on different victims. Thus, this method will be appropriate in the consideration of single rapes, for example.

However, the requirement that individual cases be mutually independent may mean that an extension to the method will be needed when serial crimes are being considered. Dependence among cases, (e.g., within litters in animal studies), can cause a phenomenon known as overdispersion. How serious a problem this will be in any future analysis of serial cases will depend, partly, on the number of serial cases there are among the totality of cases considered and on the number of crimes in each of the series. If the analysis is to concentrate on serial cases only then a check on the possible existence of overdispersion will require to be made. This is possible and there is a technique for allowing for overdispersion in the analysis. If the number of members of each of the series is small, overdispersion may not be a problem. In other crimes, not involving murder, victim statements may be available. It should then be possible to compare a victim's description of an offender with actuality. For serial cases there are multiple observations. With several different cases and offenders, verification of victim statements may be possible for certain characteristics, e.g., a study of the relationship between a victim's statements about height and the actual height of offenders may be of interest.

Although it was shown that the logistic regression model performed well with the data, an improvement in the percentage of cases detected is not the only measure of success. The contribution of the model to the time taken to solve the crime and the resources saved in its solution may be useful measures also. Further, investigating officers may wish to consider the characteristics included in the model, to see if they make practical sense or if their inclusion goes against intuition. If so, alterations to the model to take account of their concerns may be tried and tested. The models can also be continually updated as new cases arise and are solved.

It should be noted that the SAS LOGISTIC procedure uses a process known as jack-knifing when determining the values of the measures of performance, in order to eliminate the bias which can arise from testing the fitted model on the same data set as was used to construct the model originally. Each of the 320 cases was omitted in turn. A model was determined from the remaining 319 and tested on the omitted case. The final model selected is an average of the fitted models. Unfortunately, because of the confidential nature of the data it is not possible to publish details of the individual models. Requests for further details should be addressed to the Home Office Police Research Group. Further work could be done to refine the models in two ways. First, models for combinations of offender characteristics could be considered, rather than for individual characteristics in isolation, as discussed above. Secondly, models which consider so-called interactions amongst victim and crime-scene characteristics could be investigated. Thus, for example, a term $\beta_{ij} x_i x_j$ could be included in the logistic regression equation (1); such a term would take the value β_{ij} if x_i and x_j were both 1 and would take the value zero otherwise.

There are several advantages in the use of the Bayesian belief network. The relationships amongst the offender, victim and crime scene characteristics are made explicit. The structure of the graph, in terms of the nodes (characteristics) and arcs (relationships) to be included, is decided upon in discussion amongst investigating officers, statisticians and computer scientists. It is a highly interactive process and the nature of the system is such that different models can be tried very quickly.

The graphical nature of the network enables complicated relationships to be represented very clearly. Nodes which are not directly connected by an arc, in that one can only move between them along arcs by passing through one or more other nodes, are assumed to be conditionally independent of each other. This means that large complex relationships can be broken down into small simple relationships amongst only those nodes which are directly connected to each other.

The conditional probabilities relating the characteristics can be determined automatically from the database. However, they can also be obtained as measures of belief in consultation with police officers. For example, the database gives 73% of offenders with preconvictions; a senior detective suggested 70%. For marital status, the database gives 24% of offenders living with a partner; the detective also suggested 24%. For the relationship between offender and victim, the database gives 57% of offenders as being known to the victim, the detective suggested 60%. This incorporation of subjective probabilities means that the network may be extended to include characteristics which are not included in the database. Measures of belief may be used as probabilities within the system, alongside probabilities estimated from the database for other relationships. The accuracy of these measures of belief as true reflections of the uncertainty will depend to a large extent on the experience of the people providing them. Careful discussion will be required by the police officers and statisticians involved, in their determination to ensure the best values are obtained.

In theory, it should be possible to construct a network with a graph and probability tables for a particular type of crime which may be regarded as a standard network for that type of crime. The expertise of senior police officers, which would be used to construct the initial network, could then be made available to all and could be used in operational conditions.

In general, one of the great advantages of a Bayesian belief network is that it gives great flexibility to the users. The relationships amongst the characteristics are displayed in a way which may be readily understood and which has considerable intuitive appeal. If a particular network provides a relationship with which people are unhappy it is straightforward to alter the network and use the altered network for analysis. Alternatively, different networks may be constructed and their results compared. The system has much to recommend it.

The method based on Bayesian belief networks depends on considerable input from detectives for its implementation. Notice that the Bayesian belief network approach makes transparent the combination of subjectivity and objectivity in statistical analyses, something which is not always the case.

An important consideration, therefore, is the incorporation of detective expertise,and the part this plays in profiling. Holgate said that '(r)esistance to the use of statistics ... is often derived from a feeling that the practical person has knowledge that can be incorporated into the inferential process in an intuitive way, but which is not allowed for in the formal scheme of inference used by a statistician' [12]. In the models reported here, the detective expertise was concentrated in the construction of the database and in the construction of the belief networks and could be allowed for in the formal scheme of inference. The resultant statistical profiling procedure was largely a data analysis task, where the detectives provided constructive criticism of the final models. In future, investigating officers may wish to specify which characteristics (and hence their effects) should be included in which models. However, greater input of detective expertise at the modelling stage might have an effect on reproducibility, e.g., different officers may propose quite different models.

The construction of the seven-node network was carried out independently of the results from the logistic regression. The logistic regression was an entirely automatic process for the selection of the characteristics of interest. The procedure for the construction of the network as described here is a purely subjective process. In practice, both methods will be used in tandem. However, it is not unrealistic to assume that the logistic regression models could be made available in a standard package to investigating officers whilst the belief network approach could be used separately for each case to construct a network specifically for the case under investigation.

Conclusions

The potential of two statistical methods, logistic regression and Bayesian belief networks, in offender profiling has been investigated and found encouraging. The problems associated with the development of a statistical approach to specific case analysis are numerous. Possible control groups from the general population or from the total offender population based on certain crimes have been suggested, against which results from the cases have to be compared. Clear and unambiguous definitions have to be made of the characteristics to be recorded for details of the offender, the victim and the crime scene and comments about these have been made. Appropriate statistical models should be chosen. The interpretation that may be put on their results should be the subject of discussions between statisticians and detectives. The relative roles of statistician and detective in the construction of the models should be a continuing one. Continual collaboration between statisticians and detectives is required and areas for collaboration include the design of the pro forma, the collection of the data, the choice of the models, two classes of which have been discussed here, and the interpretation of the results.

Acknowledgements

Figures 1 and 2 are from Aitken CGG, Gammerman A, Zhang G, Connolly T, Bailey D, Gordon R and Oldfield R. Bayesian belief networks with an application in specific case analysis. In: Gammerman A, editor. Computational learning and probabilistic reasoning. Chichester: John Wiley & Sons Ltd., 1996: 169–184. Reprinted by kind permission of John Wiley & Sons, Ltd.

References

1. Canter D. Offender profiles. The Psychologist 1989; 2: 12–16.
2. Canter D and Heritage R. A multivariate model of sexual offence behaviour: developments in 'offender profiling'I. The Journal of Forensic Psychiatry 1990; 1: 185–212.
3. Costantino JP, Kuller LH, Perper JA and Cypess RH. An epidemiologic study of homicides in Allegheny County, Pennsylvania. American Journal of Epidemiology 1977; 106: 314–324.
4. Jason J, Flock M and Tyler CW. Epidemiologic characteristics of primary homicides in the United States. American Journal of Epidemiology 1983; 117: 419–428.
5. Riddick L, Brissie RM, Embry JH, Cumberland GD, Gilchrist TF, Glass JM and Rabren CL. Homicide in Alabama: an analysis of urban, suburban and rural murders in the deep South. Forensic Science International 1989; 40: 105–122.
6. Leonard A. The Family Background of Serial Rapists. Presented at the BPS Division of Criminological and Legal Psychology Conference, March, 1992.
7. SAS. SAS/STAT Users' Guide. Version 6, Fourth edition. SAS Institute Inc., Cary, North Carolina, U.S.A., 1989.
8. Aitken CGG and Gammerman A. Probabilistic reasoning in evidential assessment. Journal of the Forensic Science Society 1989; 29: 303–316.
9. Gammerman A. A causal probabilistic reasoning system. Fourth International Symposium on Machine Intelligence, Barcelona, Spain, 1990: 45–61.
10. Lauritzen SL and Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems (with Discussion). Journal of the Royal Statistical Society, Series B 1988; 50: 157–224.
11. Aitken CGG, Gammerman A, Zhang G, Connolly T, Bailey D, Gordon R and Oldfield R. Bayesian belief networks with an application in specific case analysis. In: Gammerman A, editor. Computational learning and probabilistic reasoning. Chichester: John Wiley & Sons Ltd., 1996: 169–184.
12. Holgate P. Contribution to Discussion of 'Statistical Inference of Phylogenies', by Felsenstein J. Journal of the Royal Statistical Society, Series A 1983; 146: 264.