Home Office

BUILDING A SAFE, JUST AND TOLERANT SOCIETY

Using homicide data to assist murder investigations

Brian Francis, Jon Barry
Russell Bowater, Nicky Miller
Keith Soothill, Elizabeth Ackerley

Home Office Online Report 26/04

Using homicide data to assist murder investigations

Brian Francis, Jon Barry, Russell Bowater,
Nicky Miller, Keith Soothill and Elizabeth Ackerley

Online Report 26/04

Acknowledgements

This study has been conducted as part of the Crime and Policing Group's programme of research into serious crime, located at the National Centre for Policing Excellence (NCPE) at Bramshill. In its efforts to respond to the operational issues facing crime investigators, the programme receives valued support and contributions from the staff of the (NCPE).

Particular thanks go to Andy Feist, for his support, encouragement and enthusiasm. We are grateful to Julian Prime and Andrew Kalinsky (Home Office) for supplying prior conviction data through the Offenders Index, and for Judith Cotton (Home Office) for the Homicide Index dataset and for fielding queries on the dataset from the research team.

The authors

Brian Francis is Professor and Director of the Centre for Applied Statistics, Lancaster University. Jon Barry is Research Associate, Centre for Applied Statistics, Lancaster University. Russell Bowater is Visiting Lecturer, Department of Statistical Science, University of Padua, Italy. Nicky Miller is a member of the Home Office Crime and Policing Group, Research Development and Statistics Directorate. Keith Soothill is Professor of Social Research, Department of Applied Social Science, Lancaster University. Elizabeth Ackerley is Research Associate, Centre for Applied Statistics, Lancaster University

Crime and Policing Group would like to thank Professor Colin Aitken for acting as the independent assessor for this report.

Contents

List of tables

List of figures

Executive summary

Background

This study explores whether routinely collected statistics on homicide can aid homicide investigation, particularly for hard-to-solve (those taking more than 28 days to solve) and unsolved cases. While most homicides are detected quickly, for up to one quarter of offences the investigative process is more complicated, and the identity of the suspect is unclear. For such cases it may be useful for the investigator to consider other sources of information to help refine lines of enquiry or establish the parameters of suspect groups, given the characteristics of the victim or the offence.

This study explored the practical application of the Homicide Index (HI) to hard-to-solve homicide investigations in predominantly adult victim homicides. The main data set consisted of 2,145 cases on the Homicide Index (covering the period between 1995 and 2000) in which an offender had been convicted of murder or a lesser offence. The HI data on offender, victim and offence characteristics were re-coded to produce a smaller number of variables. The cases were also linked to Offenders Index data on criminal convictions.

Findings

The report considers two approaches to predict the likelihood of different offender characteristics based on known characteristics of the victim and/or offence: a simple frequency approach; and a more sophisticated statistical modelling approach.

The frequency approach involved extracting cases with 'similar' victim/offence characteristics to the index case from the HI, and examining the percentage of offenders with certain characteristics (for example the percentage of offenders in certain age groups).

The statistical modelling approach involved determining which victim characteristics were important (that is statistically significant) in predicting specific offender characteristics. The victim's age and sex, the circumstances of the crime and the method of killing, were found to be significantly associated with accurately predicting the age of the offender.

The model produces results in the form a percentage score indicating the likelihood of the offender, given any combination of victim/offence characteristics, possessing a particular characteristic (e.g. the age range of the offender). Scores are provided for all categories (for instance all age ranges) so that it is possible to generate a ranked list of age bands likely to include the actual age of the offender.

Assessing the strengths and weaknesses of the two predictive approaches

Evaluating the performance of the frequency and the statistical modelling approaches revealed that, overall, the statistical modelling approach predicted offender characteristics with greater accuracy than the frequency approach. Notably, the statistical model more accurately predicted the relationship between offender and victim, the ethnic origin of the offender, and the age of the offender. However, both approaches performed with similar accuracy in predicting an offender's criminal record.

The main weakness of the frequency approach is that it requires the subjective creation of subsets of cases based on characteristics of the index case (for example female white victims, aged between 18 and 25, killed by a sharp instrument). Particular victim/offence details can be included or excluded to vary the size of the subset, which will affect the predictive quality of the analysis. Furthermore, particular problems can arise when only a small number of similar cases (or no similar cases) can be found on the HI, making results difficult to interpret operationally.

The main strength of the statistical modelling approach is that it draws on the combined power of relationships held on the database to improve the accuracy of the prediction for a given set of victim variables. Furthermore, it allows the likelihood of an offender belonging to each category within any particular characteristic, for example, each age group, to be predicted.

Although the statistical performance of the modelling approach is better *on average* than the frequency approach, there may be benefits in using both approaches in a complementary fashion in any particular case. The frequency approach is still potentially useful within the investigative context, especially when investigators want to extract a small number of similar cases from the HI in order to consider their individual profiles. Also, it is possible that the frequency approach might out-perform the statistical modelling approach for some sub-groups of homicides characterised by highly consistent offender-victim patterns. Arguably the best the way to maximise the investigative potential from the Homicide Index is to consider it as providing several complementary analytical outputs provided in a single 'package' for investigators.

Practical applications for investigators

The two methods do offer a number of potential practical applications. First, it is possible that they may help investigators consider (or challenge) the existence of alternative scenarios in a murder investigation or identify the statistical probability that a particular combination of victim and offence variables indicate a number of likely circumstances.

Secondly, and arguably of more practical value, the combined package may help investigators prioritise offender groups. For example, where intelligence-led DNA screening is being considered, it might provide an effective mechanism for prioritising age groups to be sampled by, for instance, identifying the age bands where the offenders are most likely to be present. A further potential use is in police training, helping investigating officers to better understand the complex picture of homicide, and appreciate how knowledge of different variables may alter likely outcomes.

It is important to note, however, that while the approaches advocated in this study indicate some of the potential for predicting offender characteristics, they also highlight some of the practical problems of applying complex statistical approaches to real life predictive situations.

Recommendations

This research has indicated that the Homicide Index has some potential in its application to assist ongoing, hard-to-solve (and unsolved 'cold case') homicide investigations. Consequently, the following points are recommended for consideration:

- A periodically updated copy of the Homicide Index should reside within the NCPE's Serious Crime Analysis Section (SCAS) to complement the national analytical service provided to serious crime investigations.

- A user-friendly application of the statistical model should be developed and SCAS analysts should be trained in the application and interpretation of the models. Monitoring of the application and interpretation of findings would need to be undertaken as part of the process of validation.

- A template for a 'package' of investigative advice incorporating both the frequency and statistical modelling approaches should be developed by SCAS.

- A summary of the potential for this application should be included in updated advice on the running of DNA intelligence-led screens for homicide enquiries.

1. Introduction

Background and objectives

Thankfully, homicide is a rare crime in the UK. Furthermore, the vast majority of homicides are solved and offenders are brought to justice. Unsolved cases make up a small but important minority. There were 168 cases recorded as having 'no suspect' in the period 2001/2002 (20 per cent of the 858 currently recorded homicides) (Flood-Page and Taylor, 2003). For the majority of these cases (159 cases) no suspect had been charged, with the remaining nine cases representing acquittals at court. Both types of such unsolved cases draw heavily on police resources and often attract a high media profile, which, in turn, can contribute to the fear of crime.

The Home Office Homicide Index (HI) is the primary source of official information on homicide in England and Wales. Information on all *initially recorded* homicides is recorded from information passed to the Home Office by the police. It contains information on the offence, the offender, the victim and the final outcome. Anecdotal evidence, however, suggests that relatively few detectives are aware of the HI and only a handful consult it over particular cases. According to a Home Office review of the information collected on homicide within the UK (Mayhew, 2001), the Homicide Index is routinely used to answer Parliamentary Questions on homicide and for academic research purposes. The review suggests that the Homicide Index is currently being under-utilised by practitioners and information contained within it, suitably presented, could make a contribution to the investigation of serious crime.

This study is primarily concerned with using existing information contained on the Homicide Index to help predict suspect and case characteristics from victim and offence characteristics. The use of official statistics to understand the patterns of criminal homicide is not a novel approach. For instance a similar approach was adopted by Wolfgang (1966), using official crime statistics from Philadelphia. He demonstrated strong differences between male and female victims in their relationship with the offender. Females were most likely to be murdered by a close family member (51.9% of victims), whereas males were more likely to be murdered by a close friend or acquaintance (49.7%). Although this work was not used in a police setting, he referred to the need for such research for the purposes of aiding detection.

A number of UK data sets already exist to assist the police investigation of serious crime within the United Kingdom. The CATCHEM database (Centralised Analytical Team Collating Homicide Expertise and Management) contains detailed information with respect to child homicide[1] investigations and is used to provide investigative support to detectives and behavioural investigative advisors. CATCHEM cases are initially identified from the Homicide Index and subsequently enhanced with information from case files. In addition, the NCPE's Serious Crime Analysis Section (SCAS) database provides investigative support in cases of stranger rape and homicide offenders where there is either an unknown or sexual motive and/or the homicide remains unsolved after a period of 28 days. SCAS collects data on the victim and the offence on UK cases that meet these criteria.

Other offence-offender databases have been used to provide one-off analysis of offender characteristics. For instance, Farrington and Lambert (2000) employed police recorded data from victim and witness statements to explore the usefulness of predicting offender profiles for offences of burglary and serious violence. The common approach within such studies is to examine the relation of each victim characteristic in isolation to an offender characteristic (rather than to consider the combined effect of a number of victim characteristics, a multivariate approach). Relevant multivariate work using official data is limited. Aitken *et al.,* (1996), examining child murders, was one of the first to suggest that victim characteristics could inform investigators about the nature of the offender. Karlsson (1999) used a so-called multivariate 'forensiometric' technique to predict relationship between offenders and victims, using victim and offence location information for a modest sample of 'sharp-force' homicides. No published research studies were identified, however, that appear to address the potential application of comprehensive homicide data sets to assist with *all* adult homicide investigations; they have been limited to examining subsets of the homicide population.[2] This is the purpose of the present study.

[1] A case is defined as a 'child' homicide if the victim is either 21 years of age or under for females and 17 years or under for males.

[2] The US approach which bears closest relationship to this is the work of Keppel and Weis using the Homicide Investigation and Tracking System (HITS), although this does not appear to have an explicitly predictive capacity. See for example Keppel and Weis 1993.

Objectives of the study

The project had three objectives:

- to explore, using statistical techniques, the potential contribution the Homicide Index and related data sets may provide to those investigating homicides of adults;
- to assess how the Homicide Index might best be used in a practical way to assist detectives with homicide investigations; and
- to increase the Police Service's understanding of offence-offender-victim relationships for specific sub-groups of homicides.

Methodology

This study mainly employed a multivariate statistical approach (multinomial logistic regression), using data on offence and victim characteristics supplied through the Homicide Index and the Offenders Index, to predict profiles of probable offender characteristics. Such statistical modelling techniques are useful in detecting patterns and relationships that cannot otherwise be easily identified. The offender profiles generated through this procedure give the percentage chances of the offender belonging to any particular category (for instance the chance of an offender belonging to a particular age group). A mathematical description of the technique is not provided here, but a short description can be found in the Appendix. Software to fit multinomial logistic regression models can be found in many statistical packages, including SPSS.

Data sources

Two data sources were employed for the purposes of this research: the Home Office Homicide Index and the Offenders Index. These are described below:

The Homicide Index

The main data set used in this study was the Homicide Index for England and Wales. The Homicide Index is primarily an administrative database that is used as the main source of information about the level and nature of homicide in England and Wales. It contains details of all individual offences that have initially been recorded as homicide in England and Wales. Although some of these cases are subsequently reclassified as 'no longer homicide', they are not deleted from the database. The database contains details on all suspects and all victims for each suspected homicide. Information is collected on the age, ethnicity and gender of the victims and suspects; the relationships between them; the method, circumstance/motive of the homicide; and, the final outcome of the case at court.

The computerised form of the Index dates back as far as 1977. Changes to the coding scheme were made in 1995. For the purposes of this study, cases recorded on the index between 1995 and 2000[3] were used. This was because the 1995 coding scheme provided more comprehensive information on the relationship of victim to suspect, included additional information on victim and offender ethnicity, as well as clearer coding for economic activity and occupation. Furthermore, the number of recorded murder cases from 1995 to 2000 was reasonably large (c.3,700) so the complications of combining two coding schemes (with an inevitable loss of information) could be avoided.

There are, however, several weaknesses in the way the Homicide Index records cases, particularly in the recording of *circumstance* (see Mayhew 2001 for further details). For example, *circumstances* of the offence and the *motive* of the offence are combined in the Index under a single characteristic of 'circumstance'. Thus, the category of 'arson' is classified as a circumstance of the crime, as is 'jealousy or revenge'. Furthermore, a very high proportion of homicides are categorised under the single circumstance heading of 'quarrel'. The most recent data on 'circumstances' reveal that 40 per cent of cases in 2001/2 were classified as 'quarrel, revenge or loss of temper' (Flood-Page and Taylor, 2003). While this is a decrease from the usual half or more cases classified as this in recent years, such a high proportion of the offences emerging from a rage, quarrel or revenge indicates that the 'circumstance' classification is not as insightful as might be expected. Furthermore, no information is

[3] Up to the end of March, 2000.

provided on where the homicide took place or where the victim(s) were found, while other information on the nature of the homicide (e.g. body concealment) that might have potential application to investigations is not currently recorded.[4]

With regard to completeness of the Homicide Index, Soothill *et al.* (2003) state that the Index contains nearly all those cases which were subsequently convicted of murder. They report a 99 per cent success rate in matching offenders convicted of murder on the Offenders Index between 1995 and 1997 to the Homicide Index. It is less certain, however, whether all suspected homicides are included. Furthermore, there remains the broader problem of knowing exactly what proportion of unlawful killings actually are correctly identified as such by the police (Brookman and Maguire, 2003).

The Offenders Index

A secondary source of information used in the study was the Home Office Offenders Index (OI). This is a database of all court convictions for standard list offences in England and Wales from 1963 onwards. It contains few personal details about the offender and none on the victims, but it is able to supply pre-conviction details for both offenders and victims that are not recorded on the Homicide Index.

An alternative source of information is the Police National Computer (PNC). Recent studies concerning the robustness of the Offenders Index when compared to PNC (Francis and Crosland, 2002), have shown that the two databases have approximately 70 per cent of court appearances in common (both sources of data may have missing court appearances[5]). As the Offenders Index has the advantage of allowing computer searches by type of offence as well as by name, a decision was taken to use this database rather than PNC.

Consequently, this research study explores how data contained on the Homicide Index, supplemented by additional data from the Offenders Index, can best be applied to infer the circumstances of the offence, offender characteristics and the victim-offender relationship for adult solved homicides between 1995 and 2000. First, the nature of the sample that this analysis is based upon is defined.

Defining the sample

The sample drawn upon for this study included all cases that were initially recorded as homicide that were contained on the Homicide Index between 1995 and 2000; a total of 3,684 offences. There were, however, three separate issues to be addressed in defining the sample for this study. These are discussed below.

A study of suspects or of offenders?

The Homicide Index includes cases that have yet to be brought to trial and cases where proceedings were initiated but discontinued. A decision therefore had to be taken as to whether the study sample included all cases where there was a named suspect(s), or be limited to those found guilty after the court process. It was seen as most appropriate to predict suspect characteristics for those found guilty through formal legal processes. Consequently, all cases where a suspect was not brought to trial or found not guilty have been excluded from the sample. This reduced the initial data set from 3,684 to a total of 2,145 cases. It was not possible to do anything about the relatively small number of cases that result in a conviction but are subsequently overturned on appeal.

Guilty of homicide or guilty of anything?

Most homicide suspects who are successfully prosecuted will be found guilty of murder, manslaughter or infanticide (that is, the collection of offences that are used to define homicide in some form). There will be, however, a small number of cases each year – usually between two and three per cent of the cases in recent years where there is a finding of guilt – where the suspect or suspects are found guilty of a lesser charge. Some of these convictions for lesser charges are coded specifically on the Homicide Index: grievous bodily harm (GBH), actual bodily harm (ABH) and causing death by

[4] Many of these areas are, however, being considered for collection in a redesigned HI form.
[5] OI appearances may be missing due to problems of matching criminal histories with offender details. PNC data may be missing because of weeding and records failing to be reconciled.

dangerous driving are the most common, with concealment of birth, child cruelty and aiding and abetting suicide also identified. Such outcomes occur when the court has determined that there is insufficient evidence to convict for murder or manslaughter, but there is evidence to convict for a lesser offence.

It was decided to include offences that were initially recorded as a homicide on the Homicide Index but where the offender was ultimately convicted of a lesser offence. This seems logical from an operational perspective; the study is concerned with making predictions about offender characteristics at the time of initial discovery of the homicide and relevant data on all of the suspects charged should be used in making that prediction. Table 1.1 provides a breakdown of the number of suspected homicide offenders found guilty of any offence over a five-year period (1995-2000).

Table 1.1: Number of suspected homicide offenders found guilty of an offence

	Year initially recorded					
Principal offender (a) found guilty of:	1995	1996	1997	1998	1999/ 2000	TOTAL
Murder	234	213	208	199	143	997
Manslaughter	244	220	221	245	141	1,071
Infanticide	3	4	3	7	1	18
ABH	-	1	1	1	-	3
GBH	2	2	1	-	-	5
Child cruelty	1	-	3	2	-	6
Death by reckless driving	4	5	4	2	3	18
Aiding and abetting suicide	-	-	1	1	-	2
Other lesser offence(s)	6	2	3	12	2	25
Total:	494	447	445	469	290	2,145 (b)

Note: The 1999/2000 period includes all cases recorded from January 1999 to March 2000. There were, at the time of this study (the end of 2000), a greater number of cases in progress and awaiting trial in 1999/2000, explaining the smaller number of cases in that period with a completed outcome compared to earlier years.

(a) The principal, or main, offender is the one named first in the list on the Homicide Index. See footnote 8.
(b) Some cases had more than one victim (see below).

Hard to solve cases?

A majority of homicides are solved relatively quickly. It is likely that any utilisation of the Homicide Index as an investigative tool will be on those cases that are not – so called hard-to-solve cases. It was, however, not possible to operationalise the concept of 'hard-to-solve' from information in the existing database.[6] Consequently, the analysis of the Homicide Index was not restricted to hard-to-solve cases, but included all those cases initially recorded as homicides where some finding of guilt has been made for the principal suspect (2,145 cases). This would provide a wider overview of the characteristics of all types of homicide.

Provision of criminal conviction histories

The information contained on the Homicide Index can be enhanced from other sources. For this study, it would be helpful to be able to predict whether the offender would be likely to have a prior criminal

[6] Date of arrest and date of charge would have been useful measures of 'hard-to-solve' cases; such information is not recorded in the Homicide Index. The time period between date of initial recording of the offence and date of sentencing were examined as a potential measure of 'solvability'. Factors such as delays in the court system and the difficulty of bringing a case to trial were shown, however, to affect this period, thereby not producing a satisfactory measure.

history and if so, whether there were particular types of offending (such as violence or sexual offending) present. It was therefore necessary to obtain the criminal career history for all guilty offenders. The offender's conviction history provided an additional set of outcome characteristics that could be predicted by the statistical analysis.

Furthermore, in most cases, the police are usually quick in determining whether the victim has a criminal history. The victim's conviction history could be a helpful addition to the dataset because it offers additional *explanatory* characteristics that might be used to predict offender characteristics.

The names, dates of birth and gender of 2,570 of the 2,573 guilty offenders in the sample of 2,145 cases were submitted to the Offenders Index to obtain their criminal conviction history.[7] All guilty offenders on the Homicide Index should have an entry on the Offenders Index, and so, in theory, the criminal histories of all offenders should be present. Matching, however, was only moderately successful, with 2,310 names matched out of the total number of 2,570 names supplied, leaving 10 per cent (260) that remained untraced on the Offenders Index. In terms of principal suspects,[8] an entry on the Offenders Index was found for the principal suspects in 1,940 of the 2,145 cases. A further manual matching exercise resulted in the identification of an additional 92 principal suspects, including two of the three who had a missing date of birth and so were not sent for the automatic matching process. This left 113 (5%) principal offenders for whom no satisfactory match could be found on the Offenders Index, although the Homicide Index indicated a conviction should be present.[9]

Where the data were available, the names, dates of birth and gender of the victims of the 2,145 cases on the Homicide Index where the principal suspect was found guilty were also submitted to the Offenders Index to identify their criminal histories. The reasons for doing this were twofold. Firstly, this is a task that would be carried out by an investigating officer, and such information would be routinely available to help guide investigators. Secondly, it was expected that information on whether the victim had a criminal history might be informative in terms of understanding the offender victim/relationships, circumstances etc. In the 2,145 cases in the sample there are 2,223 victims, but 396 (18%) of these could not be sent for matching to the Offenders Index: about half (201) were victims aged under ten, and so would not have any official criminal record available in any case; the date of birth was missing for the remaining 195 victims, preventing identification. Of the 1,827 victim names which *were* submitted to the Offenders Index, a criminal record was traced for 705 (39%), with 1,122 victims remaining untraced. Among these 705 were the principal victims in 698 of the 2,145 homicide cases under consideration (33%).

Unlike the suspect matching exercise, there is no information for victims on the Homicide Index which indicates that a criminal record on the Offender Index should be found. While a proportion of the victims were successfully traced on the Offenders Index, it is likely that there are others which have been missed. The *suspect* matching had a 90 per cent success rate – if the rate were similar for the *victim* tracing, 784 (instead of 705) of the submitted 1,827 victims should have criminal records (43%). A similar proportion with criminal records among the 195 victims above the age of criminal responsibility (aged ten years or more), who were not passed on to the OI because of missing data, would result in another 84 victims with criminal histories. None of the 205 victims under the age of ten would, of course, be found on the Offenders Index. It is therefore estimated that, in the 2,145 homicide cases where the principal suspect was found guilty, the actual number of victims having a criminal record is close to around 868 (39% of all the victims in these cases, or 43% of victims aged ten or over).

Summary of sampling procedure

In summary then, between 1995 and March 2000[10] the Homicide Index contained a total of 3,684 cases that were initially recorded as homicide, involving 3,845 victims and 4,441 suspects. Of these

[7] Three convicted suspects could not be submitted to the Offenders Index for matching due to their date of birth – a required field for the matching process – being missing.

[8] The principal, or main, suspect is the one named first in the list on the Homicide Index. The order in which the suspects are placed is determined operationally when the returns are completed by Police Forces, as is the case with the victims. However, the initial order of the suspects can change – for example, a suspect later convicted of murder will be placed above one convicted of manslaughter, who will in turn move above an acquitted suspect.

[9] This factor needs to be considered when carrying out statistical analyses involving the criminal history of offenders.

[10] Due to the change in recording practices in the *Criminal Statistics: England & Wales*, whereby figures since April 1998 have been reported for financial rather than calendar years, the 1999 period includes January to March of 2000.

cases, 3,576 (97.1%) were recorded as single victim homicides, and the rest multiple victim homicides, with seven being the maximum number of victims in any one case. Only three-quarters of the cases (2,752) had a single suspect, with a maximum of 11 suspects among those with more than one suspect, and 298 (8.1%) cases have 'no suspect' recorded at all.

Of the initial 3,684 cases, there were 2,145 where a conviction against the principal suspect was obtained. This conviction ranged from murder, manslaughter or infanticide, through to ABH, GBH, child cruelty, causing death by reckless driving, aiding or abetting suicide, to 'other lesser offences'.[11] While in total there were 2,223 victims and 2,573 suspects where a conviction had been obtained in these 2,145 cases, the series to be analysed consists of details relating only to the *principal* victims and *principal* suspects – 2,145 of each.

Structure of the report

The report is organised into four further chapters. Chapter 2 describes in more detail the nature of the sample that forms the basis for this research, focusing on the characteristics of the victims, offenders and the homicides under consideration. Chapter 3 details the development of a series of statistical models that enable offender characteristics to be predicted from available victim characteristics. Chapter 4 demonstrates the practical application of the model to a number of cases recorded on the Homicide Index. Chapter 5 reviews the main findings of the research and summarises the recommendations arising from the study.

[11] There were no convictions for 'concealment of birth' – the other named offence – in the Homicide Index data between 1995 and March 2000.

2. A summary description of the cases

This section describes the basic information on offenders and victims for the sample used for this study. It describes the key variables used as well as setting the scene for the later sections which describe the process of prediction.

The Homicide Index data were re-categorised into a series of broad themes. All unsolved homicide investigations are characterised by the presence and absence of certain pieces of information. Some pieces of information will routinely be known to the police early on in an investigation (e.g. basic victim characteristics such as age and gender), others are likely to become known as the enquiry progresses; and some may not be known even after an offender is apprehended (precise motive, circumstance).

Victim characteristics, offender characteristics and 'offender profiles'

A *victim characteristic* is a piece of information that the police would often know quite soon after a body had been discovered. Victim characteristics can relate to the status of the victim such as their age, sex, ethnicity, job status and criminal history; other victim characteristics can relate to the status of the body when it is found. Operationally, investigating officers will usually know the gender and ethnic group of the victim relatively quickly within an investigation; forensic science can usually determine such information even if the body is badly decomposed or incomplete. For many cases, the pathologist will provide information on cause of death to the investigating officers. Once the body has been identified information on the precise age, job status and so on will be established.

An *offender characteristic* is a piece of information that would generally only be known when the homicide has been 'solved', although eyewitness and physical evidence may provide information that could indicate the sex of the offender. Furthermore, DNA evidence (if available) may allow certain individuals to be eliminated from the enquiry. Some offender characteristics relate to the characteristics of the main or principal offender such as their age, sex, ethnicity, criminal history and the offender's relationship or social connection with the victim (e.g. stranger or acquaintance). Other offender characteristics can relate to the other people that may have been involved in the killing such as the number of offenders.

One characteristic that was difficult to classify was the circumstance or motive for the killing. As noted, the 'circumstance' and the 'motive' are recorded under a single variable on the Homicide Index. The difference can be described as follows: the circumstance of the killing relates to the sequence of events that led to a killing. The motive of the crime relates to the aim or intention of the offender in carrying out the offence and is less easy to determine; it may, in fact, never be known. For example, the circumstance of the crime may involve a sexual component, but the motive for the killing may not be sexual. Conversely, the offence may be sexually motivated but there may be no obvious sign of sexual activity. The crime may be staged by the killer to mislead the investigating officers; there are many killings where there appears to be no motive.[12] Motive can, therefore, be extremely difficult to establish conclusively, particularly at the early stages of an investigation. Due to the fact that in some instances the circumstances and motive will be known but in others it would not be, 'circumstance' was treated as both a victim characteristic (if thought to be known) and an offender characteristic (if thought to be unknown or uncertain) in the analysis.

For the purposes of this study, an *offender profile* was defined as a set of offender characteristics with their associated percentage chances of occurring. The aim of the approach was to *estimate* an offender profile based on known victim characteristics. Knowledge that there is a high chance that the offender belongs to a particular age range, for example, may be helpful in setting suspect parameters, prioritising searches, or otherwise providing investigators with alternative scenarios in hard-to-solve cases.

The Homicide Index, supplemented with Offenders Index data, contains information that can act as useful victim (explanatory) or offender (outcome) characteristics. Each of these characteristics was categorised, enabling the estimated percentage chance for any category in the offender profile to be

[12] Over 16 per cent of the 2145 cases under consideration in this study were recorded as having no motive.

calculated. So, for example, the offender's age was categorised into seven groups, and the percentage chance of the principal offender being between 21 and 24 years can be estimated for any set of victim characteristics.

The *full data set* comprised information relating to offender characteristics of age, sex, ethnic origin, circumstances, relation to victim and method of killing (a total of 2,145 cases). The data set containing offender characteristics of prior criminal record, prior violent offence, prior drugs offence and prior sexual offence was referred to as the *criminal career data set*. There were 113 cases for which the offenders' prior criminal records were not available. Consequently, the size of this data set was reduced to 2,032 cases.

Table 2.1 lists the victim and offender characteristics extracted from the Homicide Index and Offenders Index. Each of these is discussed in more detail below. Descriptions of the necessary coding changes are given in Appendix A.

Table 2.1: Offender and victim characteristics used in the analysis

Victim characteristics	Offender characteristics
Age	Age
Sex	Sex
Ethnic origin	Ethnic origin
Circumstance	Circumstance
Prior criminal record	Prior criminal record
Prior violent conviction	Prior violent conviction
Prior sexual conviction	Prior sexual conviction
Prior drug conviction	Prior drug conviction
Job status	Relationship
Method	Number of offenders

The characteristics of the dataset

The following section outlines the main characteristics of the combined Homicide Index/Offenders Index dataset. For several key headings, variables were combined to allow the dataset to be more manageable for undertaking the statistical analysis. These are highlighted as appropriate.

As Table 2.2 illustrates, the vast majority (97.7%) of these 2,145 cases involved a single victim, with only 50 cases having more than one victim (between two and seven) (Table 2.2). In contrast, one in five cases involved more than one offender (between two and 11), although not all of these individuals were convicted. There were 2,223 victims in total.

Table 2.2: Number of victims and offenders per case

	Victims		Offenders	
	Number of cases	%	Number of cases	%
One	2095	97.7	1716	80.0
More than one	50	2.3	429	20.0
Total	2,145	100.0	2,145	100.0

The gender of the principal victims and principal offenders is given in Table 2.3. While females account for around a third (31.9%) of *victims*, only one in ten (10.3%) of the *offenders* were female.

Table 2.3: Sex of principal victims and principal offenders

	Victims		Offenders		England and Wales population
	N	%	N	%	%
Male	1,461	68.1	1923	89.7	48.6
Female	684	31.9	222	10.3	51.4
Total	2,145	100.0	2,145	100.0	100.0

Note: Population figures from Office for National Statistics: mid-1999 Population Estimates: England and Wales; estimated resident population, revised in light of the results of the 2001 Census

The ethnic background of the principal victims and principal offenders was more consistent (table 2.4). The proportions in each category were quite similar, with just over three-quarters of both principal victims and offenders being classified as White (77.8% and 78.4% respectively).

Table 2.4: Ethnicity of principal victims and principal offenders

	Victims		Offenders		England and Wales population (a)
	N	%	N	%	%
White	1,669	77.8	1,682	78.4	91.3
Black	158	7.4	195	9.1	2.8
Asian (Indian sub-continent)	101	4.7	105	4.9	4.7
Other	44	2.1	52	2.4	..1.2
Not recorded/not known	173	8.1	111	5.2	0.0
TOTAL	2,145	100.0	2,145	100.0	100.0

(a) Taken from Census 2001 Table S104.

Age profiles for victims and offenders were quite dissimilar. Victims covered the full age range with sizeable numbers in the extremes. Offenders were mostly concentrated into a narrower age range, with a high proportion in their 20s and early 30s. In addition, as the sample contained only those offenders with a conviction, this restricted the lower age limit for offenders to ten years (being the age of criminal responsibility). In the upper age range, the oldest *offender* was 78 years, while the oldest *victim* was 98 years, and only 40 (1.9%) of offenders were aged 60 years or over, compared with 261 (12.2%) of victims. Such differences require the use of different age bands for victims and offenders; the age bands used in the statistical analysis that follows are shown in Table 2.5.

Table 2.5: Age of principal victims and principal offenders

Victims			England and Wales population 1999		Offenders			England and Wales population 1999
	N	**%**	**%**			**N**	**%**	**%**
0 – 2	135	6.3	3.7		**Under 18**	151	7.0	22.9
3 – 13	52	2.4	14.3		**18-20**	228	10.6	3.7
14 – 17	88	4.1	4.9		**21-24**	361	16.8	4.6
18 – 24	332	15.5	8.3		**25-29**	393	18.3	7.0
25 – 29	260	12.1	7.0		**30-39**	597	27.8	15.5
30 – 39	468	21.8	15.5		**40-49**	233	10.9	13.2
40 – 49	339	15.8	13.2		**50+**	182	8.5	33.1
50 – 59	210	9.8	12.2					
60+	261	12.2	20.9					

Note: Population figures from Office for National Statistics: mid-1999 Population Estimates: England and Wales; estimated resident population, revised in light of the results of the 2001 Census

The Homicide Index includes two variables relating to the victim's job status: *economic position* and *key occupation*. Following the conventions used in previous studies which have used the Homicide Index (see Soothill *et al*. 1999), these were combined to give a single *job status* measure having the following six categories; 'manual', 'non-manual', 'student', 'retired', 'inactive' and 'other'. No information is currently collected through the Homicide Index on the job status of offender. Appendix A explains how the original variables were combined, while Table 2.6 shows the numbers falling into each of the new, broadly defined categories. A significant proportion of the victims were not currently in employment, with just under one-tenth (9.7%) retired and one-third (34.1%) classified as unemployed or otherwise economically inactive.

Table 2.6: Occupation of principal victims

	N	%
Manual	339	15.8
Non-manual	172	8.0
Student	105	4.9
Retired	209	9.7
Inactive	731	34.1
Other	589	27.5
TOTAL	2,145	100.0

Table 2.7 gives details of the victims that could be identified as having a prior criminal record. For 190 victims it was not possible to obtain sufficient information from the Homicide Index to search the Offenders Index. Furthermore, there were 174 victims under the age of criminal responsibility (i.e. ten years), who would not have any recorded criminal history on the Offenders Index. These two types of victim were combined into a single category for the statistical analysis. The remaining victims who *were* sent for matching were classified as either having a previous criminal record or not, according to whether their details were found on the Offenders Index. Offenders' criminal histories are considered separately towards the end of this chapter.

Table 2.7: Existence of criminal records of principal victim

	N	%
Existence of criminal record	698	32.5
No record found on OI*	1083	50.5
Unknown/under 10 years old	364	17.0
TOTAL	2,145	100.0

Note: A failure to locate a record on the OI did not necessarily imply that the victim did not have a criminal record. There may have been only convictions which were not 'standard list', and therefore not eligible for inclusion on the Offenders Index. Alternatively, there may have been a small but critical difference between the personal details held on the Homicide Index and those on the Offenders Index, resulting in no match being found (see Chapter 1 for more details).

More detailed information on the *type* of criminal history can be extracted from the Offenders Index. Arguably the most useful operational measure would be the offence for which the offender had been previously convicted. Convictions for three offence types were analysed in detail – violent offences, sexual offences and drugs offences. Previous research (Soothill *et al*. 2002) has shown that being convicted of a violent or drugs offence affects the risk of the offender committing murder later in an offender's criminal career, but there is little large-scale research on the criminal histories of victims of homicide. Table 2.8 shows the number of victims who had previously been convicted in one of these offence groups (a victim can be in the table more than once) Appendix B contains details of which offences are contained in each group.

Table 2.8: Nature of offending history of principal victim (a)

	N	%
Contains a violent offence	377	17.6
Contains a sexual offence	53	2.5
Contains a drugs offence	152	7.1

(a) A single victim could have offences in more than one category and therefore could appear up to three times in the table.

The relationship of the victim to the offender is recorded in some detail on the Homicide Index. It was necessary, however, to collapse some of these categories to produce a more manageable measure for use in the statistical analyses. Six categories were constructed (see Appendix A for what is covered in each). Table 2.9 provides a frequency and percentage breakdown of the relationship of the principal victim to the principal offender.

Table 2.9: Relationship of principal victim to principal offender

	N	%
Offspring	153	7.1
Spouse/lover	453	21.1
Other relative	156	7.3
Acquaintance	771	35.9
Stranger	387	18.0
Unknown	225	10.6
TOTAL	2,145	100.0

The majority of homicides on the database involved someone known to the victim (71.4%), with less than one in five being committed by a stranger (18%). A further one in ten (10.6%) homicides had insufficient information recorded for a judgement to be made on the relationship between the victim to the offender.

The Homicide Index records the *method of killing* in one of 19 distinct categories, several of which are rare and consequently contain very few cases. Re-coding of method into a more manageable eight categories was undertaken to improve the analytical process (see Soothill *et al*. 1999 for details and Appendix A for more information on the combination of categories). The method by which the victim was killed could not be determined in only a small number of cases (1.9%), while the use of a sharp instrument was the most common method of killing and used in just under two-fifths of cases (37.9%) (Table 2.10).

Table 2.10: Method of homicide

	N	%
Sharp instrument	814	37.9
Hitting/kicking	279	13.0
Blunt instrument	263	12.3
Strangulation, asphyxiation or drowning	253	11.8
Shooting	101	4.7
Fire	38	1.8
Other	356	16.6
Not known	41	1.9
TOTAL	2,145	100.0

Issues around the classification of circumstances in which homicide cases occur have already been touched on in Chapter 1. The Homicide Index has a total of 22 categories; these were collapsed into the eight groups shown in Table 2.11. Unlike the method of the homicide, which can often be determined even when there is no suspect, the circumstances may or may not be clarified until a suspect is discovered. This was reflected in the fact that 347 (16%) of the sample cases were classified as having insufficient information to determine the circumstances. Where the circumstances *were* determined (84%), there was a heavy bias towards the 'rage or quarrel' category (54%). This pattern was unsurprising, as it is frequently found across a range of subsets of homicide (Soothill *et al*. 1999).

Table 2.11: Circumstances of the homicide

	N	%
Rage or quarrel	1158	54.0
Jealousy or revenge	120	5.6
Sexual	54	2.5
Theft or other gain	158	7.4
Feud	16	0.7
Irrational act	84	3.9
Other circumstances	208	9.7
Unknown	347	16.2
TOTAL	2,145	100.0

Offenders' pre-convictions

Tables 2.2 to 2.11 have shown the profile for the 2,145 principal offenders and victims that made up the sample for this study. Problems were encountered with the matching of offenders in the Homicide Index sample with the Offenders Index. This resulted in 113 principal offenders for whom no adequate match could be found on the Offenders Index (in spite of the Homicide Index recording a conviction). While this does not affect any of the victim characteristics, or how other elements of the offender are dealt with, it was necessary to omit the 113 cases from the statistical analysis *when the focus was on predicting the criminal record of the offenders*. Thus for the analysis of any offender criminal record characteristics, the series was reduced to those 2,032 cases where a conviction against the principal suspect was obtained *and* the criminal record of the principal offender was obtained from the Offenders Index.

As with the victims' criminal record characteristics, four measures were calculated for analysis. The first measure was whether or not the offender had any convictions *prior* to the conviction date associated with the Homicide Index case. The remaining measures were whether or not the offender had prior convictions for violent offences, for sexual offences and for drugs-related offences (Appendix B contains details of which offences were contained in each group).

Table 2.12 shows the numbers of the 2,032 principal offenders who have prior convictions of any type, or of the three particular types considered (an offender can appear in the any of the bottom three rows of the table more than once).

Table 2.12: Presence of types of previous criminal records of offenders

	N (2,032)	%
No pre-convictions	568	28.0
Any pre-convictions	1,464	72.0
Total	2,032	100.0
Type of pre-conviction:		
Violent	899	44.2
Sexual	107	5.3
Drugs	318	15.6

Although we need to be careful in making comparisons due to the differential attrition of cases for victims and offenders, several points are worth making. Compared with the victims (Table 2.7), where approximately a third (27%) had a previous criminal record, approximately three-quarters (72%) of offenders had at least one conviction prior to their homicide (or lesser charge). The proportion of offenders with previous convictions for the three particular types of offence were all more than double the proportions for victims, with previous violent offences being especially common (44% of offenders had a pre-conviction for violence compared with only 18% of vicitms).

This figure of 72 per cent for this sample of offenders is marginally higher than that quoted by Soothill *et al.* (2002) in reporting on a sample of murderers. In that sample, 68 per cent had a prior conviction. However, it is worth pointing out that the present sample is defined by an offence which is initially recorded as homicide; it therefore will include as offenders those eventually convicted of manslaughter and other less serious offences. Thus the two samples are not strictly comparable.

The next Chapter looks at ways in which information on the Homicide Index could be used to develop investigative advice to Senior Investigating Officers (SIOs).

3. Using the Homicide Index for development of investigative advice

The key objective of this research was to assess how the Homicide Index could best be used in a practical way to assist detectives with homicide investigations. Useful applications would include the provision of behavioural investigative advice (formerly referred to as 'offender profiling' advice) and for analytical support, both of which inform the development of investigative strategies.

Behavioural investigative advice and analysis are predominantly requested for serious sexual offences and homicide where the offender is unknown. For cases of serious sexual assault and rape, the victim is usually able to provide information concerning the offender such as their relationship with them, an indication of their ethnicity, whether more than one offender was involved, and an estimate of the offender's age. With homicide cases, such information is frequently not readily available to the senior investigating officer (SIO). This kind of information is usually helpful in focusing the direction of investigations (e.g. setting suspect parameters for searches or developing lines of enquiry).

One source of data that is increasingly drawn upon by both the behavioural investigative advisors (BIAs) and crime analysts towards the construction of their advice is the SCAS database at the NCPE[13]. The homicides on the SCAS database are not, however, representative of all homicides that occur within the UK – it is much more geared towards hard-to-solve or sexually motivated homicides[14]. Because of its restricted criteria for cases, it does not, for instance, provide information on domestic homicides which account for a substantial proportion of all cases (Brookman and Maguire, 2003).

Predicting offender characteristics from victim characteristics

There are two approaches to using a database such as the Homicide Index to predict offender characteristics. These can be summarised as the *frequency* approach and the *statistical modelling* approach. Each is summarised briefly below, with examples to illustrate.

The raw data/frequency approach to guiding investigations

The simplest way to use the raw data to gain information about the likely characteristics of offenders is to tabulate the frequencies for particular combinations of offender characteristics. Some examples of these tabulations were shown in the previous Chapter. For example, from Table 5, we know that 78.4 per cent of principal offenders in all homicide cases resulting in a conviction of any kind, were White, 9.1 per cent were Black, 4.9 per cent were Asian and 2.4 per cent were recorded as Other. These tabulations could be used in a predictive manner. So, using the example above, we could predict the chance of the principal offender being white as 78.4 per cent regardless of the specific details of the victim. We have termed this the *frequency approach* - it uses cross tabulations of the raw data only on the sample of offenders. The frequency approach provides a baseline against which other approaches can be compared.

Additional victim characteristics can be used to refine these predictions. For example, consider the problem of predicting the likely age of the principal offender given the age of the principal victim. That is, given the age category that the victim belongs to, what is the chance that the offender will lie in a particular age category? A simple solution to this problem is to cross-tabulate the raw data. So, for each victim age category, the percentage falling within each of the categories of the offender age can be tabulated, thereby obtaining an offender age profile for each victim age. These are displayed below in Table 3.1.

The offender age profiles for victim age categories are given on the rows of the table. For example, if the victim is between 30 and 39 years, there is a 13 per cent chance that the offender will be between 21 and 24 years; if the victim was between three and 13 years, there would be a four per cent chance that the offender was aged 50 years or over.

[13] The database is based upon the Violent Crime Linkage Analysis System (ViCLAS) developed by Canada. It stores the details of relevant offences including details of the offence itself, including locations; victimology; offender details (if known); verbal behaviour; and, behavioural and details of forensic information, if available.

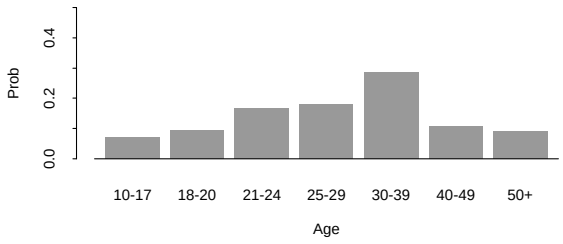
[14] Information is collected on the SCAS database for those murder offences where the motive is known (or believed) to be sexual; the motive is unknown; and, the offence remains undetected after 28 days.

Table 3.1: Offender age profile as a percentage, given age of homicide victim based on raw data

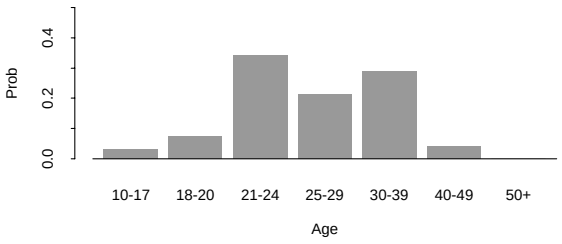
Age of principal victim	Age of principal offender							
	10-17	18-20	21-24	25-29	30-39	40-49	≥50	all ages
0-2	4	11	34	21	26	4	0	100
3-13	8	12	19	15	38	4	4	100
14-17	33	24	14	7	14	6	3	100
18-24	10	16	29	20	17	5	2	100
25-29	5	8	18	30	30	7	4	100
30-39	3	10	13	19	39	11	4	100
40-49	5	7	10	13	32	21	12	100
50-59	5	10	12	16	24	14	20	100
>=60	10	8	12	14	22	13	21	100
All ages	7	11	17	18	28	11	9	100

Figures 3.1 – 3.10: Age profiles of the offender, given the age of the victim.

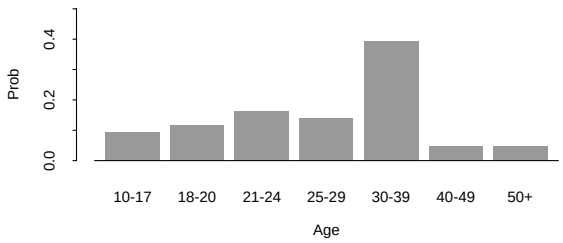
Victim Age Values: NONE



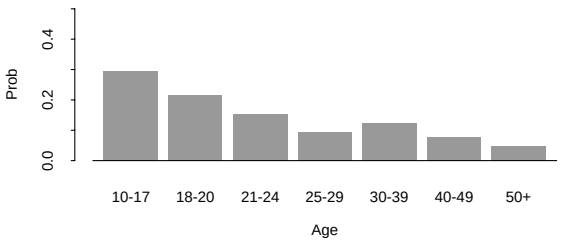
0-2



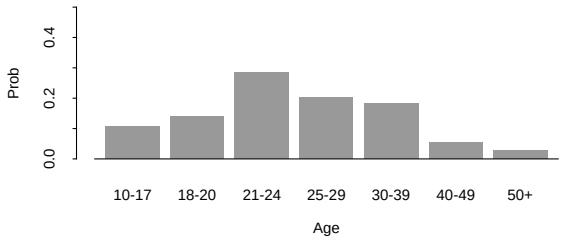
3-13



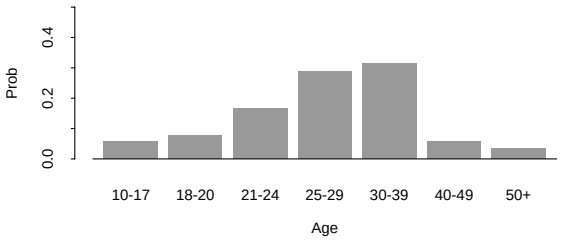
14-17



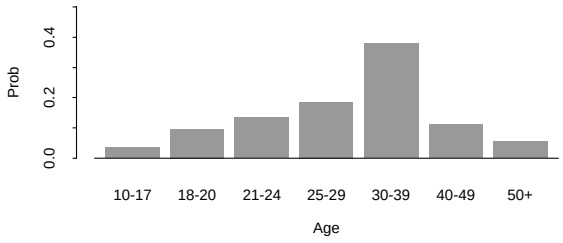
18-24



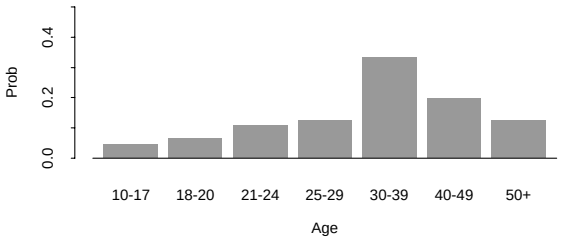
25-29



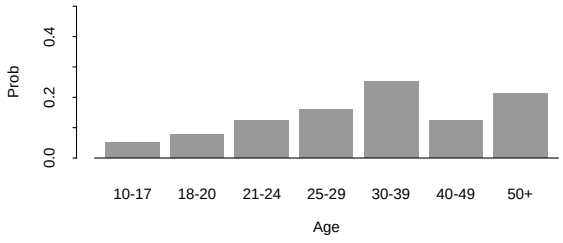
30-39



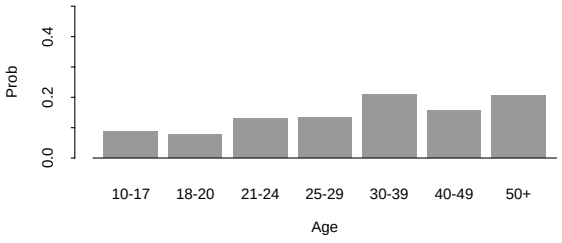
40-49



50-59



60+



Figures 3.1 to 3.10 present this approach graphically through a series of bar charts. The first bar chart, labelled 'Victim Age Values: none' (Figure 3.1) reflects the frequency of the different suspect age ranges for *all victims*. When the victim's age is unknown, the most common age group for the offender in the sample is 30-39 years. Figure 3.2 gives the likely distribution of offender age categories when it is known that the victim is under the age of two years. The graph shows that most offenders lie between 21 to 24 years. The remaining bar charts (Figure 3.3 to 3.10) provide similar offender frequencies for different age groups of victims. Thus, we can see the changes to the offender age probability profile, as the age of the victim becomes older. The last chart (Figure 3.10), for example, illustrates that victims over the age of 60 years become more likely to be killed by older offenders (i.e. over 50 years).

In reality, of course, it is likely that SIOs are usually very rapidly aware of multiple characteristics of the victim (for example, their sex and their age). The use of the frequency approach when there is information on two victim characteristics (sex and age) is illustrated in Figures 3.11 to 3.22. For male victims aged 60 and over and female victims aged 60 and over, the charts show offender profiles generated around four offender characteristics:

- the age of the offender;
- the sex of the offender;
- the offender's relationship to the victim; and,
- offender ethnicity

The first column (Figures 3.11 to 3.14) presents the offender age distribution using data from all victims (that is assuming that nothing is known about the victim). Figures 3.15 to 3.18 show the profile obtained when we have a *male* victim over 60 years, and Figures 3.19 to 3.22, the profile for when the *victim* is a female over 60 years.

A clear difference exists between males and females in terms of offender age. For elderly male victims, the most common offender age category is 30 to 39 years (Figure 3.15), although the likelihood of being a young offender (0 to 17 years) or an older offender (over 40 years) both increase compared to all victims (Figure 3.1). In contrast, for elderly female victims, the most common age of the offender is above 50 years. There was little difference between male and female elderly victims when the profile for the sex of the offender was examined. However, the age distributions for victims do show differences in terms of the victim-offender relationship. For both male and female elderly victims, offences committed by acquaintances are less common than for all victims. For males, the most likely relationship is a stranger crime, whereas for females it is hard to discriminate between a relation, an acquaintance and a stranger. For both male and female elderly victims, the chance of being killed by a husband, wife or lover is lower compared with all victims.

Problems with using the frequency approach

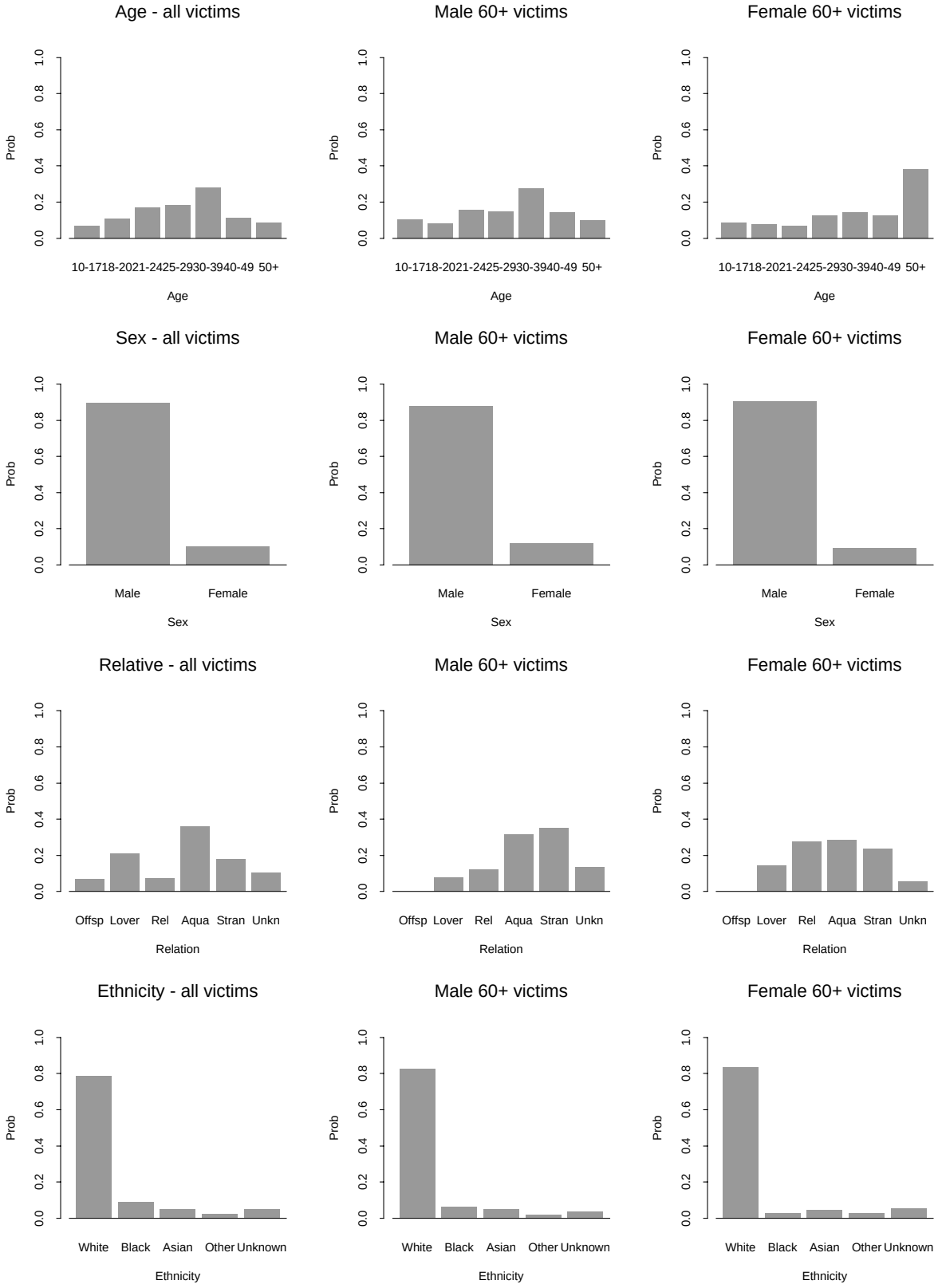
Using more than two victim characteristics can provide further refinements to the analysis. For example, take the case where there is a need to identify the likely offender characteristics for an unemployed female victim aged 34 years, stabbed with a sharp instrument. Using the frequency approach, all cases where the principal victim was a female around the age of 34 years; where the recorded method was stabbing; and, where the victim was unemployed, would need to be selected and the related offender characteristics observed.

While this might be viewed as a simple and obvious approach to prediction, it has limitations. If we actually examine the database, there were two cases found with the above characteristics. Both were male offenders, aged 47 in the first case and 50 in the second; both were ex-lovers of the women and for both the circumstances were "Rage or quarrel", and both offenders were white. The agreement in the profiles for these two cases could well lead investigators to limit their search to white males between 47 and 50. However, if we change the age of the woman to be 35 rather than 34, then the age range for the six cases found widens to be from 24 to 50 – and the victim/suspect relationships range from spouse, common-law husband through to other acquaintance. It is clear that while insight is gained, a method of averaging over similar cases is needed, and this suggests the need for a statistical modelling approach.

The likelihood of 'small numbers' of cases occurring on the Homicide Index when examining relatively small subsets of victims with specific characteristics inevitably increases; this in turn can exacerbate the problem of interpreting any findings from the frequency approach. In such instances, judgements

may need to be made as to the relative importance given to particular victim variables (for instance certain victim variables could be ignored to increase the number of cases studied). Furthermore, the interpretation of results can prove difficult. This is likely to be the case except where the results yield highly consistent patterns (e.g. for the selected victim variables it ends up that all are offenders male, etc.). Likewise, combinations of victim characteristics that yield no cases can be problematic. In essence, while by no means valueless to investigators, the effectiveness of the raw data will in part depend on the skill and experience of the individual manipulating the data.

Figures 3.11 to 3.22: Offender profiles for offender age, sex, relationship to victim and ethnicity, for all victims, male elderly victims and female elderly victims



Although the frequency approach can be insightful and is a good first step in predicting offender characteristics, statistical modelling (i.e. use of sophisticated statistical models to predict offender characteristics) will provide predictions by detecting important patterns in the data between the victim characteristics and the offender characteristics. Its advantage is that it will provide predictions for all combinations of victim characteristics, even where there are no prior cases in the database. This approach also provides a way of dealing with a large number of victim variables, where it is probable that no exact match to the characteristics of a case would be found in the database. It is thus a reliable and consistent way to predict offender profiles for any combination of victim variables. Modelling such data needs considerable computer power, but once the models have been estimated, the results can be displayed at any time. Easy-to-use software could be written to display these results for police investigations.

The statistical modelling approach

Statistical models identify the nature and strength of the relationship between an offender characteristic and each victim characteristic. The relationships are summarised in a mathematical formula, which can then be used to make predictions. The skill of statistical modelling is thus to find a model which will represent all of the major relationships within the data set, without representing the chance fluctuations which always occur.

A range of statistical models could be used. The model that was considered most appropriate for the Homicide Index data was the Multinomial Logistic model. This was because the offender characteristics were mainly (or could be easily re-coded as) categorical, and most possessed more than two categories. Full details of how to fit this model using the statistical modelling package GLIM are provided in Aitkin *et al.* (1989); Aitkin and Francis (1992); and, Francis *et al.* (1992). Other software can also be used (R, SPSS, STATA, S-PLUS).

Chapter 2 summarises the ten offender characteristics that represented the statistical profile of the offender, and the ten victim characteristics that were used in the statistical modelling to predict the offender profiles. The statistical modelling considered each offender characteristic separately. The aim of the process was to determine the victim characteristics that were important in influencing the offender profile for that characteristic. For each offender characteristic, the statistical model was built up stage by stage. At each stage, the victim characteristic which best improved the prediction of the observed offender data was added to the model. This process was continued until all the important (that is, statistically significant) victim characteristics had been included.[15]

The victim characteristics used to predict each offender characteristic are listed in Table 3.2. For example, to predict the number of offenders involved in a homicide, the victim variables of age, sex, ethnic origin, circumstance and method were required. In contrast, to predict the likelihood of the offender having a prior sexual offence in their criminal history, information on the method used in the murder was needed. It is worth pointing out here that whether the characteristic 'circumstance' is to be treated as a victim characteristic or as an offender characteristic will depend on the homicide under investigation. For example, the circumstance of the homicide may be evident, in which case it can be treated as an additional explanatory victim variable. Self-evidently when modelling 'circumstance' as an offender characteristic, the model clearly cannot also use 'circumstance' as a victim characteristic.

Note that the notation Age*Sex in the table represents a statistical interaction between the victim characteristics Age and Sex and appears in many of the final models. This interaction allows the model to take into account the fact that the victim's age will affect offender profiles in a different way for males and females. For example, Age*Sex is important when predicting the age of the offender; this means that the relationship of offender age to victim age is different for male and female victims.

[15] Each explanatory victim characteristic was fitted separately and the degrees of freedom and deviance for that model was calculated (see Aitkin *et al*., 1989 for technical details). Characteristics were considered to be important only if they increased the fit of the model by a statistically significant amount as measured by the reduction in deviance compared to 95th percentile of the appropriate chi-squared distribution. Because of the complexity of the models and the amount of computer time needed to fit them, only major interaction effects were considered. For all characteristics apart from offender's ethnicity, the only interaction term considered was between victim's age and sex. For offender's ethnicity, interactions between victim's ethnicity and method and between victim's ethnicity and criminal record were tried – but these proved not to be statistically significant.

Table 3.2: Important victim (explanatory) variables for each offender (outcome) variable

Offender characteristic	Important victim variables
Age	Age, sex, age*sex, circumstance, method
Sex	Age, sex, age*sex, circumstance, method
Ethnic origin	Ethnic origin
Circumstances	Age, sex, ethnic origin, job status method
Prior criminal record	Age, circumstance, criminal record, prior violent offence, prior sexual offence, prior drugs offence, job status, method
Prior violent offence	Age, sex, age*sex, circumstances, criminal record, job status
Prior sexual offence	Method
Prior drug offence	Age, sex, age*sex, circumstances, criminal record, job status
Relationship	Age, sex, age*sex, ethnic origin, circumstances criminal record, job status, method
Number of offenders	Age, sex, ethnic origin, circumstances, method

A measure of model performance

The purpose of the statistical modelling approach was to estimate profiles for each offender characteristic given a set of victim characteristics. To evaluate the performance of the models, the estimated profiles were compared with the characteristic of the *actual* offender. Clearly, for the model to perform well the actual offender characteristic should be predicted with a high degree of chance. To measure the performance of the models, the following procedure was applied.

The full data set of 2,145 cases was randomly split into two data sets; one contained 1,609 cases and the other 536 cases. These represented 75 per cent and 25 per cent of the cases respectively.[16] The larger subset was used in the modelling procedure to determine the important victim variables and to determine the model of how these victim variables were related to the relevant offender variable. This data set was labelled the *modelling* data set. The smaller data set of 536 cases, referred to as the *validation* data set, was used to test how well the resulting models performed.[17]

For each of the offender characteristics, the statistical model outlined in Table 15 was used to estimate the offender profiles within the validation sample. Thus, for example, when estimating the sex of the offender, the chance that the offender was male, and the chance that the offender was female, was calculated for each case. This then allowed the prediction to be compared with the actual value of the offender characteristic.[18]

This validation procedure provides an obvious summary measure of how well the model performed. This measure is referred to as 'L' and measures the likelihood that something is the case. For instance, the model was applied to seven cases in the validation data set to determine how well it

[16] A similar procedure was carried out for the *criminal careers* data set, where the two files were of size 508 and 1524 respectively. A uniform random number between 0 and 1 was generated for each data point; a number less than 0.75 meant that the observation was placed in the modelling subset; a number greater than or equal to 0.75 placed the observation in the validation subset.

[17] These 536 new cases were not used in the estimation of the offender profiles and therefore provided a good assessment of the quality of the technique.

[18] The same procedure was also applied to the criminal career data set.

predicted the offender's age. The model calculated the 'L' score, that is the *likelihood*[19] of the offender belonging in each of the age bands. These results are presented in Table 3.3.

Table 3.3: Predicted chances of actual offender ages for seven validation cases

Case	Age band actual age lies within	Predicted chance of that age band from model 'L' (%)
149	30-39	45
519	21-24	42
527	30-39	33
297	40-49	19
66	10-17	15
132	18-20	2
433	40-49	3

The actual age of the offender in Case 149 was between 30 and 39 years; the model predicted this with a 45 per cent chance of occurrence. The actual age of the offender in Case 519 was between 21 and 24 years; the model predicted this with a 42 per cent chance of being the case. The model does not merely identify the category with the highest chance, it also predicts chances for all other categories. So, for example, for Case 149 the model predicts while there is a 45 per cent chance of the offender being aged 30-39 years, there is a 27 per cent chance that the offender is aged between 40-49, and a chance of two per cent of the most unlikely age group of 10-17. This, critically, allows for suspect groups to be ranked by particular characteristics (in this case, age) and may have clear application for prioritised screening of offender sub-groups.

As Table 16 makes clear, by no means all the validation case age bands were predicted with a high probability by the model. For Cases 433 and 132, for example, the model predicted that there were only three per cent and two per cent chances that the offender would be in the age bands in which he or she was actually located. However, it should be emphasised that a low chance is not a zero chance – a three per cent chance for age group 40-49 (case 433) means that for three cases in every hundred with the victim characteristics of case 433, the principal offender's age is estimated to be in age group 40-49, and case 433 is simply one of those cases.

Evaluating the performance of the different approaches

In this section the *overall performance* of the statistical modelling approach described above is compared with the *overall performance* of the frequency method. With this second approach, the frequency of the characteristics of the set of offenders was used to calculate the profiles for the dataset as whole. So, for example, Table 6 revealed that, from the full dataset, 89.7 per cent of offenders were male. This is the probability (or likelihood) we would attach to any future offender being male (note that in the analysis below we exclude cases in the *modelling* data set when calculating these probabilities so that our validation cases are entirely separate from the cases used to calculate the profiles).

We have used a summary statistic, $\overline{\overline{L}}$ which is the geometric mean[20] of all such probabilities over all cases in the validation dataset, expressed as a percentage. So how do we interpret $\overline{\overline{L}}$? It is an attractive measure to use in that it has a direct meaning – it is the geometric average of the actual outcome chances over the cases in the validation dataset. We want $\overline{\overline{L}}$ to be as high as possible so

[19] Statistically, this is the likelihood of the estimated model (produced from the modelling dataset) given the data in the validation dataset.

[20] Statistically, it is the average log-likelihood (per case) of observing the validation sample given the parameters estimated from the modelling sample, expressed as a likelihood. If there are N cases in the validation sample with probabilities p_1, p_2, …, p_N, then $\overline{\overline{L}} = (p_1 \times p_2 \times \ldots \times p_N)^{(1/N)}$ is the geometric mean of the probabilities.

that we predict offender characteristics well. In the extreme, $\overline{\overline{L}}$ = 100 per cent is the best possible outcome and indicates that the model has predicted the actual category perfectly for all of the cases in the set. However, such an ideal value will not be achieved in practice. Conversely, a value of $\overline{\overline{L}}$ close to 0 per cent suggests that the model has predicted the actual categories very badly, with very low chances. $\overline{\overline{L}}$ will depend on the number of categories in the offender characteristic of interest – lower values of $\overline{\overline{L}}$ will be obtained where there are more categories. We will primarily look for percentage changes in $\overline{\overline{L}}$ in moving from one method to another.

The first column of Table 17 shows the value of $\overline{\overline{L}}$ obtained using the simple frequency approach outlined above. The second column shows $\overline{\overline{L}}$ for each of the offender characteristics fitted using the statistical modelling approach (multinomial logistic model), expressed in the table as percentages. This model uses the specified victim information given in Table 3.2. The overall values of $\overline{\overline{L}}$ produced by the statistical model are mostly all higher than their corresponding value obtained using the frequency approach (although not for the variable Sexual Offence). The improved performance in the prediction of offender victim relationship (a 52 per cent improvement) was particularly noticeable. The equally high scores for $\overline{\overline{L}}$ when predicting prior convictions for a sexual offence needs to be carefully interpreted. This reflects the fact that since the majority of offenders did not have a previous conviction for a prior sexual offence, in close to eight cases out of ten, both approaches accurately predicted this to be the case.

Overall therefore, this analysis demonstrated that the statistical modelling approach, on average, predicted the actual offender characteristics better than the frequency approach. In other words, the statistical models were, overall, shown to be successful.

Table 3.4: Comparison of the frequency and modelling approaches using $\overline{\overline{L}}$ (a)

Offender Variable	Frequency $\overline{\overline{L}}$ (%)	Model $\overline{\overline{L}}$ (%)	% improvement
Age	15.6	17.4	11
Sex	68.3	73.6	8
Ethnic origin	45.2	54.7	21
Circumstance	23.6	27.8	18
Criminal record	54.7	56.1	3
Violent offence	50.5	51.1	1
Sexual offence	77.47	77.48	0
Drug offence	64.3	64.7	1
Relation	19.5	29.6	52
Number of suspects	61.6	65.5	6

(a) Based on the validation datasets only.

The third column shows the percentage improvement of $\overline{\overline{L}}$ for the statistical modelling approach compared with the frequency approach (with the frequency value of $\overline{\overline{L}}$ used as a base for the percentage). The statistical model performed better for offender relationship, ethnicity and circumstance (an improvement of 52%, 21% and 18% respectively). However, the improvement for the four characteristics indicating the offender's criminal record and past offending were all small, with prediction of the offender's prior sexual offence showing no improvement at all.

This approach inevitably paints the frequency method in its most unsophisticated form. It simply does not allow any refinement of the raw dataset to allow for sensible sub-group analysis on smaller groups

of victim characteristics, (e.g. homicides of female victims aged 30-39, strangled). It is quite possible that for some sub-groups of victims, particularly those with strong or consistent offender characteristics, the frequency approach may actually outperform the statistical modelling approach. The difference between them is that the statistical modelling approach will utilise information from a wide range of cases in building the model – the frequency approach will look only at a subset of cases. Nevertheless, Table 3.4 confirms that the victim variables used in the statistical model were providing important information to help predict suspect characteristics and were, overall, an improvement on the calculation of the offender profiles over those derived from the frequency approach.

4. Practical application of the statistical modelling approach

To illustrate the model described in Chapter 3, seven cases were selected from the validation dataset (three to illustrate the prediction of the main offender characteristics and four to illustrate the prediction of the criminal careers characteristics) to illustrate the predicted profiles of the offender for given victim characteristics. The victim and offender characteristics for these cases are shown in Tables 4.1 and 4.2.

These cases were chosen on the basis of how well the model performed in making predictions, choosing both well performing and poorly performing cases. Figure 4.1 (case 149) provides an example of how well the model works in practice; Figure 4.2 (case 66) demonstrates an average performance of the model in predicting offender characteristics; and, Figure 4.3 (case 433) shows how it may not always perform as effectively. It is important to be aware that the model will not always perform effectively as it bases the predictions on probabilities. Consequently, there will be cases where the predicted probability of an actual offender characteristic is low. Such cases will naturally occur from time to time.

The predicted offender profiles for the main offender characteristics are shown in Figures 4.1 to 4.3 as a series of bar charts. For each offender characteristic, the outline predicted by the statistical model is given. In addition, the actual observed value for the offender is highlighted in grey. Similar charts are shown for the criminal career characteristics in Figures 4.4 to 4.7.

We illustrate how to read these charts by examining Figure 4.2 (case 66). This was an average performer and showed some strengths and limitations of the statistical modelling approach. The victim was an 18-24 year old Asian male, who was unemployed, and stabbed in a rage. The top left graphic shows the predicted age profile of the offender. The prediction shows that the most likely age group for the offender was 21-24 years (with a probability of approximately 25 per cent). The other age groups, however, also possess similar probabilities. For example, there was a 20 per cent chance that the offender could be between 25-29 years; a probability of 17 per cent they could be aged between 18-20 years; and, a 15 per cent chance that the offender could be ten-17 years. In fact, as Figure 4.2 demonstrates, the actual age of the offender fell in the ten-17 year age group (the shaded bar); the model predicted that such cases would occur 15 per cent of the time for the given victim profile.

Exploring the other graphics in Figure 4.2, it can be seen that the chance of the offender being male was high at approximately 90 per cent; the chance of an acquaintance killing was also relatively high, at about 60 per cent. Indeed, the actual offender was a male acquaintance. The model also predicted a chance of a single offender at 70 per cent and multiple offenders at about 30 per cent. While most cases involve single offenders, a small but sizeable minority of cases involves multiple offenders, and in this example this was the case. The circumstance of the offence was predicted to be a 'rage' offence with a chance of 70 per cent, and indeed, the offence was classified as such. Finally, the ethnic group was predicted. The most likely group was given as Asian with a probability of about 55 per cent, and the next highest, white at 30 per cent. In this case, the offender was indeed Asian.

It is important to examine the *whole* offender profile looking at the categories with small chances as well as those with large. For example, in looking at the relationship prediction within Figure 4.2, the likelihood of the victim being an offspring or lover/spouse of the offender was relatively small. It was also unlikely that the victim was a stranger to the victim. For ethnic group, it can be seen that the offender was unlikely to be black. However, *unlikely is not the same as impossible*, and there will be a small proportion of cases with this victim profile; where the offender will be a child of the offender or where the offender is black. It is important for BIAs, crime analysts and SIOs to be aware of this when developing investigative advice and strategies. The model should *not* be used in isolation but taken in conjunction with other advice/evidence provided by those involved within the investigation; the model is there to *assist* the SIO in parameter setting; it is not a substitute for it.

Table 4.1: Victim and offender characteristics for three selected cases (personal and crime offender variables – full dataset)

	Excellent performer	Average performer	Poor performer
Case ID	149	66	433
Victim variables:			
Age	30-39	18-24	40-49
Sex	Female	Male	Male
Method	Knife etc.	Knife etc.	Blunt instrument
Circumstance	Rage/ quarrel	Rage/ quarrel	Theft
Job status	Manual	Inactive	Non-manual
Ethnicity	White	Asian	White
Criminal record	Unknown	Yes	Yes
Offender variables:			
Age	30-39	10-17	40-49
Sex	Male	Male	Female
Ethnicity	White	Asian	White
Relationship of victim to offender	Spouse/ lover	Acquaintance	Spouse/ lover
Number of suspects	1	>1	1
Circumstance	rage/ quarrel	Rage/ quarrel	Theft

Table 4.2: Victim and offender characteristics for four selected cases (offenders criminal career characteristics – criminal careers dataset)

	Good performer	Good performer	Average performer	Average performer
Case ID	24	100	115	421
Victim variables:				
Age	40-49	50-59	0-2	25-29
Sex	Male	Male	Male	Female
Method	Fire	Blunt instrument	Other	Hit/ kick
Circumstance	Rage/ quarrel	Insane	Rage/ quarrel	Insane
Job status	Inactive	Non-manual	Other	Non-manual
Criminal record	Yes	No	Under 10	No
Violence Offence	Yes	No	No	No
Sexual offence	No	No	No	No
Drugs offence	No	No	No	No
Offender variables:				
Criminal record	Yes	No	Yes	Yes
Violent offence	No	No	Yes	Yes
Sexual offence	No	No	No	Yes
Drugs offence	Yes	No	No	Yes

Figure 4.1: Predicted personal and crime offender profiles for case 149

Victim: 30-39 white female, manual worker, unknown criminal record, stabbed in rage/quarrel

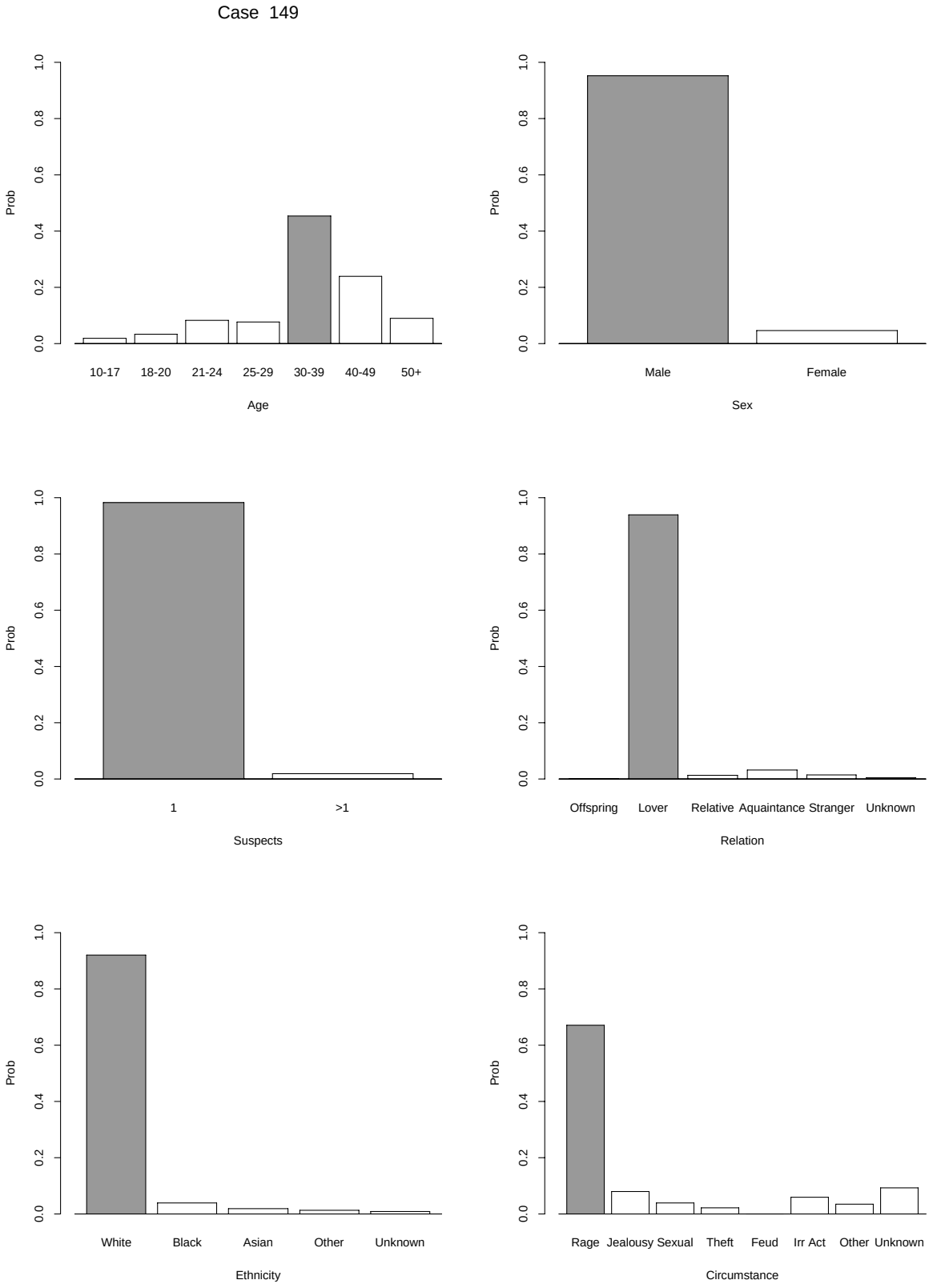


Figure 4.2: Predicted personal and crime offender profiles for case 66

Victim: 18-24 unemployed Asian male, previous criminal record, stabbed in rage/quarrel.

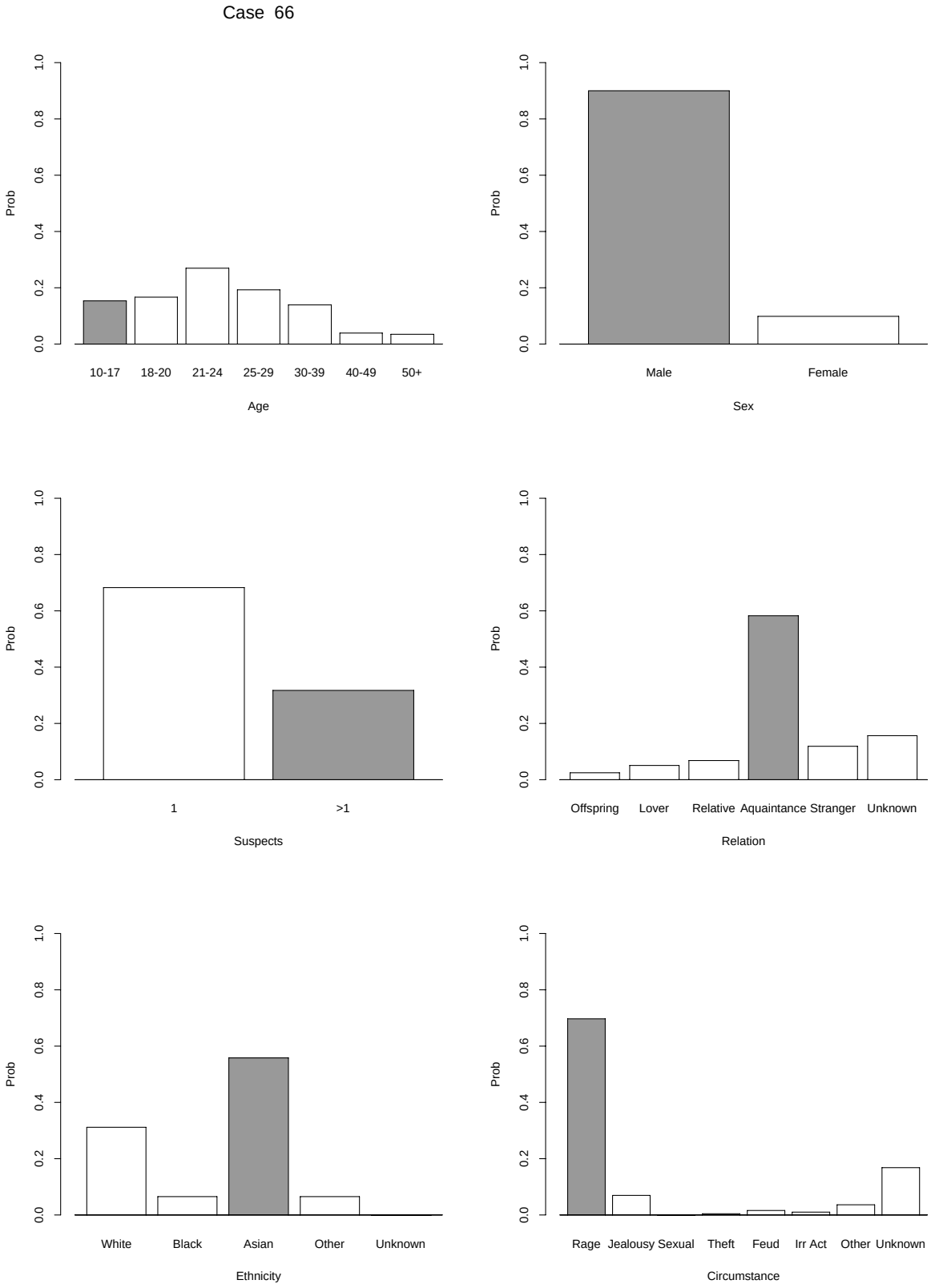
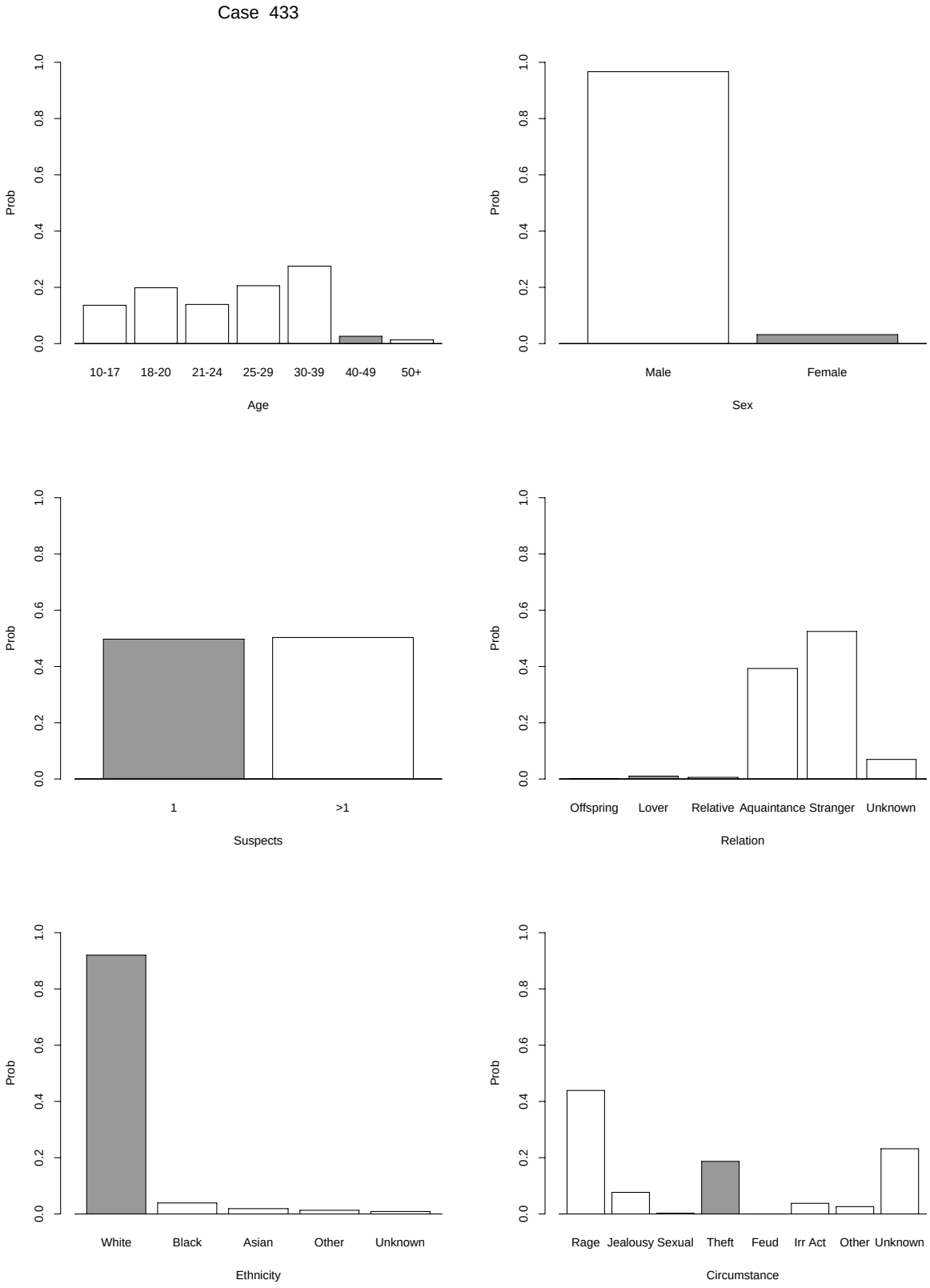


Figure 4.3: Predicted personal and crime offender profiles for case 433

Victim: 40-49 white male, non-manual occupation, previous criminal record, killed by blunt instrument, theft motive



The predicted offender profiles for the criminal career offender characteristics are shown in Figures 4.4 to 4.7. They give the profile predicted by the statistical model for each of the four criminal career characteristics (whether the offender has a criminal record or a prior drugs, violence or sex conviction). The actual observed value for the offender was again highlighted in grey.

Figures 4.4 and 4.5 can be examined in more detail to explain the interpretation of these graphs. Figure 4.4 shows the prediction for an unemployed male with a previous violent offence killed by fire-setting or arson as a result of in a rage or quarrel. In this case, the model predicted that the offender was very likely to have a criminal record, with a reasonably high chance of a prior violent offence but with a low chance of prior drug and sexual convictions. Indeed, the actual offender did have prior convictions for violence but not for a sexual or drugs offence.

In contrast, Figure 4.5 shows the prediction where the victim was a male non-manual worker, with no criminal record, killed by a blunt instrument by someone of imbalanced mind. Here, it appeared to be very unlikely that the offender would have a prior criminal conviction. The chances of a prior drugs or violent conviction were also both low. The chance of a prior sexual conviction, although very low, was slightly higher than the previous example. The actual offender had no prior criminal record. These two examples are ones where the model's prediction worked well; Figures 4.6 and 4.7 illustrate additional examples that were more typical.

Figure 4.4: Predicted criminal career offender profiles for case 24

Victim: 40-49, unemployed male, previous violent offence, killed by fire in rage/quarrel.

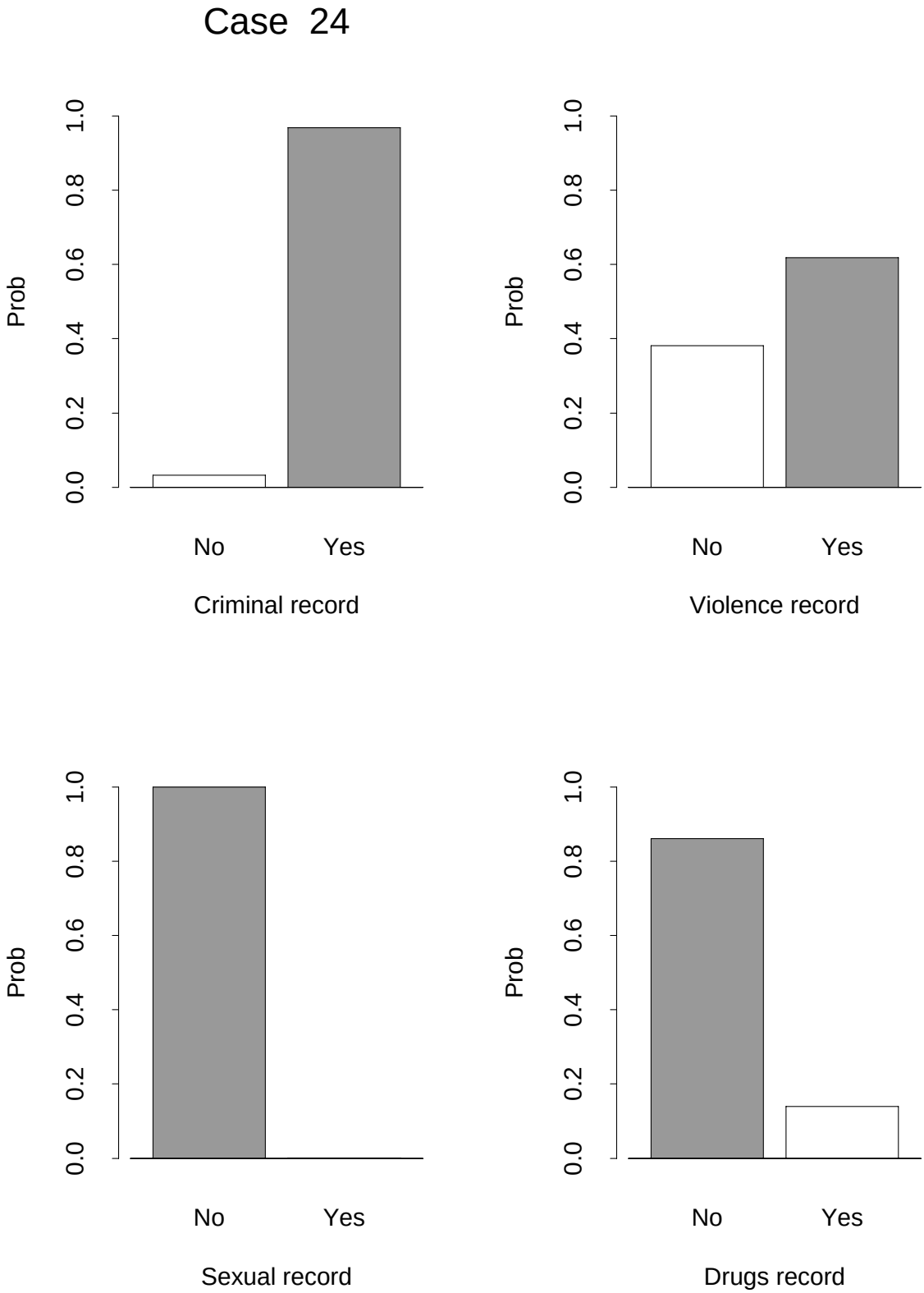


Figure 4.5: Predicted criminal career offender profiles for case 100

Victim: 50-59, male non-manual worker, no previous criminal record, killed by blunt instrument apparently by someone of imbalanced mind.

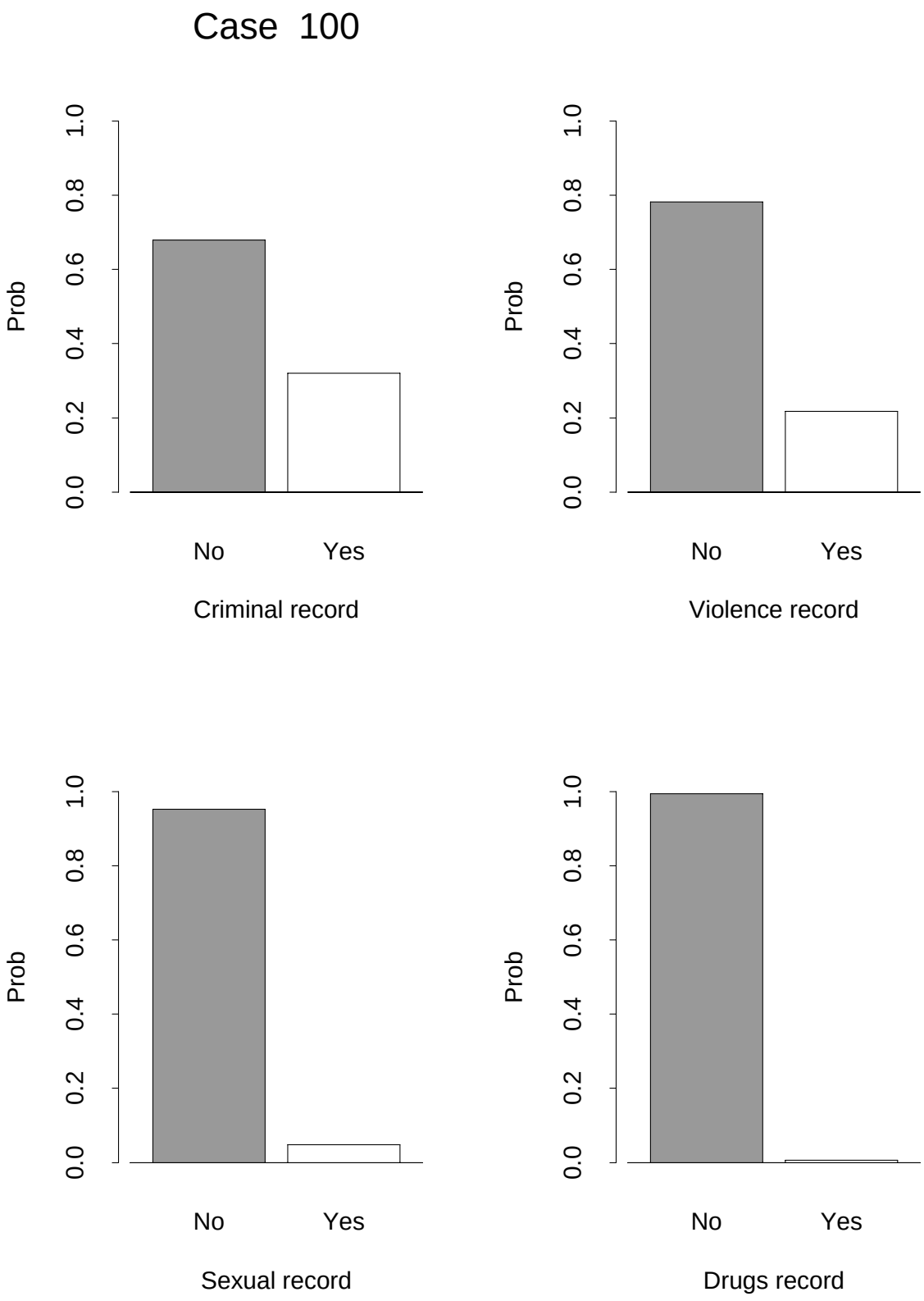


Figure 4.6: Predicted criminal career offender profiles for case 115

Victim: 0-2 male baby, killed in rage/quarrel.

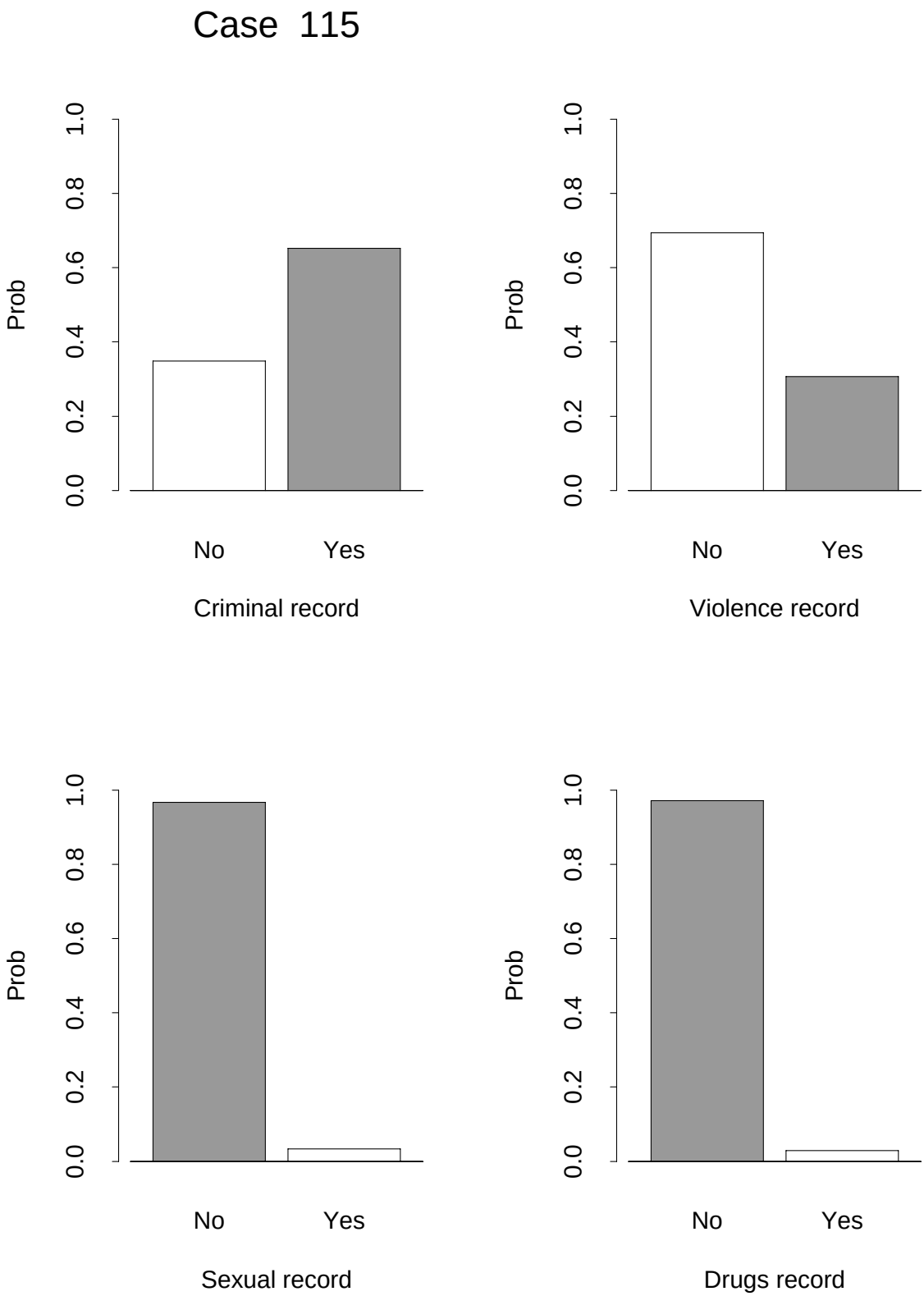
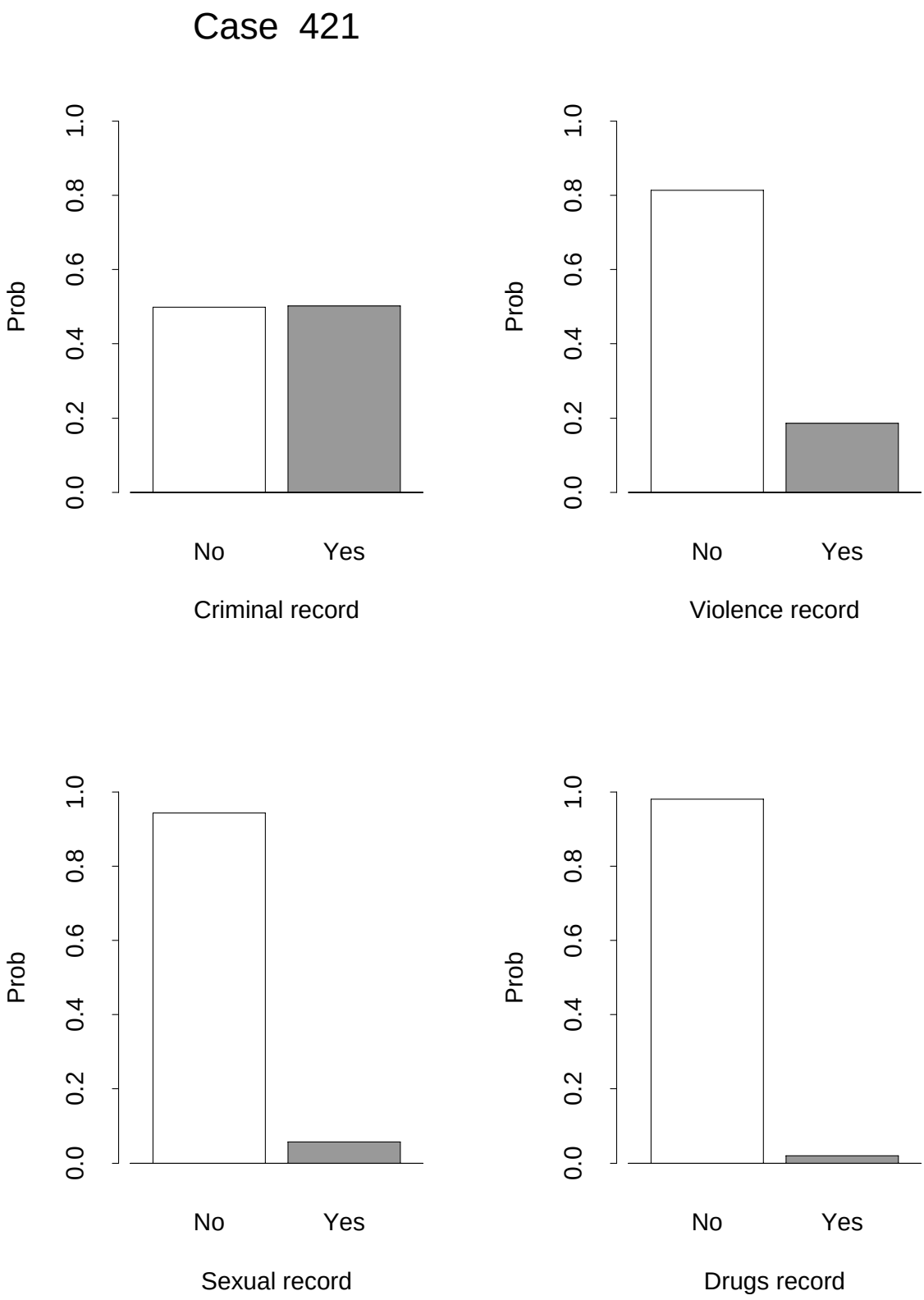


Figure 4.7: Predicted criminal career offender profiles for case 421

Victim: 25-29 female, non-manual worker, no previous criminal record, hit/kicked to death apparently by insane person



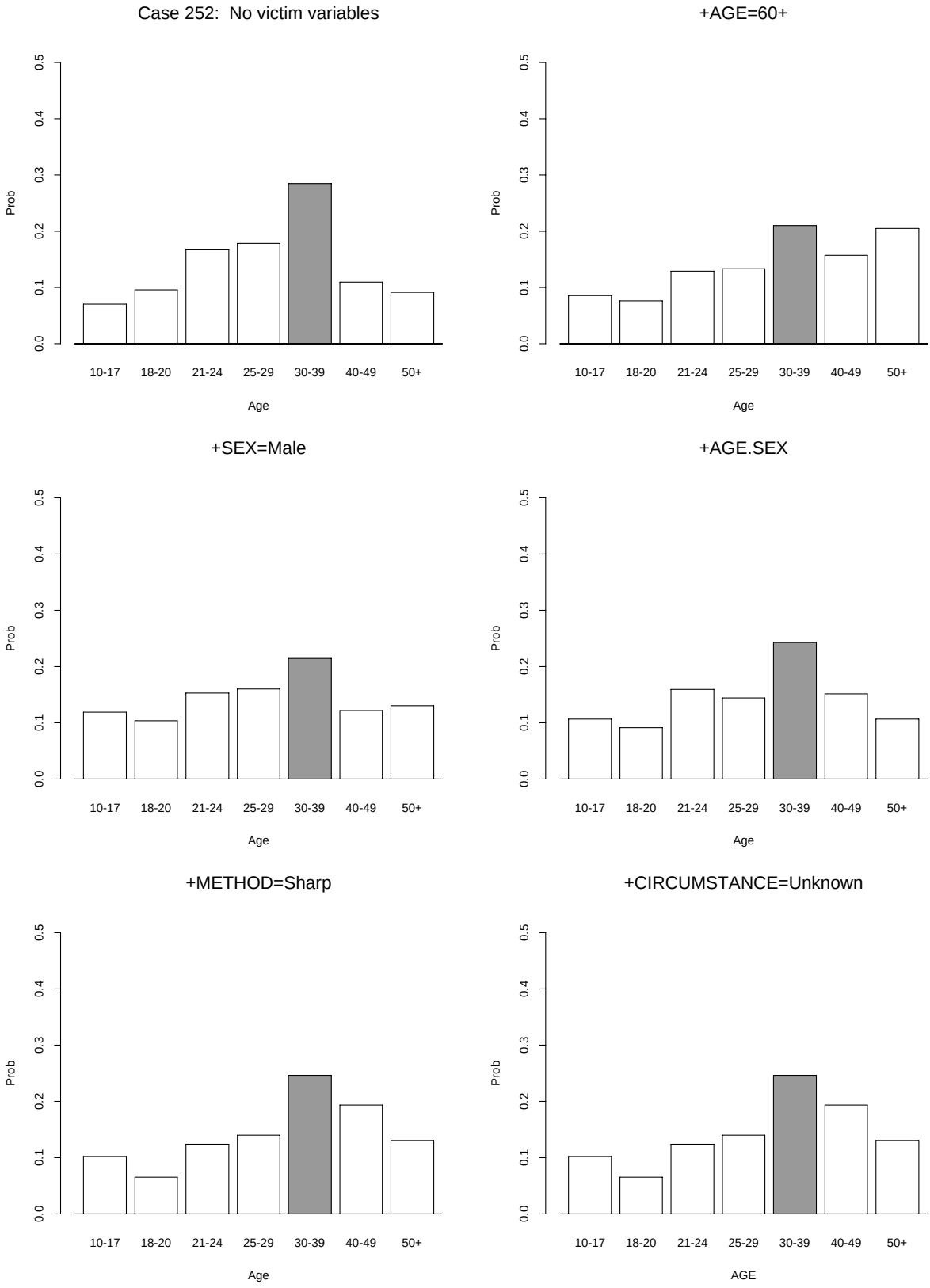
Building an offender profile step by step

The case examples demonstrate that the offender profiles vary substantially between different types of victim. It is useful to see how an offender profile is built up as each piece of victim information is added. This would reflect the investigative process in reality, as it is often the case that different pieces of information may enter an enquiry at different times. As additional pieces of information concerning the victim and/or the homicide enter the investigation, the profile of the offender may change. Figures 4.8 to 4.13 illustrate how this process could be applied operationally, building an offender profile for age. Case 252 was chosen at random to illustrate this process.

Figure 4.8 shows the predicted age profile when nothing was known about the victim. Adding in the piece of information that the victim was aged over 60 years, Figure 4.9 demonstrates that the offender age profile changes; the chance of the offender being aged over 50 years increases, and the chance of the offender being aged 30-39 years declines. Adding in the piece of information that the victim was male places an additional two terms in the statistical model: Sex and the Age*Sex interaction. The offender age profile has changed again; the chance of the offender being over 50 years has declined somewhat, but is still larger than that for Figure 4.8. Finally, adding in the fact that the victim was stabbed, and that the circumstance was unknown produces the profile shown in Figure 4.13. The age profile has again changed, with the profiles of the last two offender age groups increasing slightly.

Of course, in most circumstances, the victim variables are usually known and available to the investigator at an early stage of an enquiry. This example has illustrated how the offender profile can be adjusted as any additional victim information is fed into the model. It clearly demonstrates that the process of developing offender profiles using the statistical modelling approach can be dynamic.

Figures 4.8 to 4.13: Building an offender age profile for case 252 step by step



5. Conclusions and recommendations

While most homicides are detected quickly, there is a small proportion where the investigative process is more complicated, and the identity of the suspect is unclear. For such cases it may be useful for the investigator to consider other sources of information to help refine lines of enquiry or establish the parameters of suspect groups, given the characteristics of the victim or the offence.

This study has attempted to explore the practical application of a primarily administrative database, the Homicide Index, to actual hard-to-solve homicide investigations in predominantly adult victim homicides. By using reduced re-coded Homicide Index variables, and combining the Homicide Index data with other information on the offender and the victim's criminal careers, it is possible to build up a picture of a offender's likely characteristics using different statistical techniques. The report considers two approaches: a simple frequency approach; and a more sophisticated statistical modelling approach to predict the likelihood of different offender characteristics.

The frequency technique involves trawling the Homicide Index to create a subset of cases with similar characteristics to that for the victim/offence in a case under investigation. The main weakness of the frequency technique is that it requires the subjective creation of subsets of cases. Particular victim variables can be included or excluded to vary the size of the subset (and so ease the reading of the data). Excluding discriminatory victim variables such as job status or ethnicity may help in increasing the size of the subgroup datasets but the impact on the predictive quality will be unclear. Indeed, how effective this is in producing meaningful results will depend on the abilities of the individual performing the task. Furthermore, even if it is possible to select relevant variables skillfully, this can still generate findings that are difficult to interpret operationally, particularly if based on small numbers of cases.

The more complex statistical modelling approaches relies on the detection of complex patterns and relationships between all relevant offender and victim variables held on the database. In simple terms this means that each of a range of dependent variables related to the victim and the crime scene are used to help predict the likely characteristics of an offender. The principal advantage of the modelling approach is that it draws on the combined power of relationships held on the dataset to improve the accuracy of the prediction for a given set of victim variables. It also allows additional knowledge from the investigation to be added into the model as and when it is available to the investigation. A measure of the statistical model's performance has been included that suggests that, overall, the modelling approach yields more accurate results when compared to the overall performance of the frequency approach. The improvement is most marked in predicting the relationship between the offender and suspect, the ethnic origin of the offender, and the age of the offender. Moreover the modelling approach predicts profiles for each offender characteristic and allows SIOs to determine the likelihood of an offender belonging to all categories within any particular characteristic (for instance all categories of relationship and all age ranges).

At this stage several strong notes of caution need to be sounded. First, the modelling procedure predicts a *distribution* or *profile* of possible values, and the chances of the offender belonging to each category of the profile. It is invalid to interpret such a profile by looking at that category with the highest chance, and taking that to be the prediction. Secondly, while the statistical performance of the 'modelling approach' is *on average,* better than that of the frequency approach, there would appear to be merit in applying both approaches to any particular case in a complementary fashion.

The frequency approach may well out-perform the statistical model for some sub-groups of homicides that are characterised by highly consistent offender/victim patterns – this will depend on the complexity of the model. Furthermore the frequency approach may be particularly helpful to investigators if they want to extract a small number of similar cases from the Homicide Index and consider their overall profile. However, problems can arise if no or very few similar cases can be found.

In summary therefore, the way to maximise the investigative potential from the Homicide Index is to consider it as providing several complementary analytical outputs provided in a single 'package' for investigators. Ultimately, this might consist of three discrete pieces of analysis:

- listings of similar cases meeting narrowly drawn victim-based criteria using the detailed HI codes, providing case and offender characteristics;

- a cross tabulation of offender characteristics using broader HI categories and linked with the OI data on criminal careers; and
- the outputs of the statistical model.

The skill of the analyst would be to consider the value of all three outputs in a way that maximised the potential value of investigative advice, given what else was known to a particular enquiry.

Notwithstanding some of these cautionary observations, the methods presented offer a number of potential practical applications. First, it is possible that they may help investigators consider (or challenge) the existence of alternative scenarios in a murder investigation or identify the statistical probability that a particular combination of victim and offence variables indicate a number of likely circumstances. Some hard-to-solve homicides present SIOs with alternative and contradictory circumstances/offence scenarios. Here, investigators need to begin to make well-founded judgements about the relative strength of one scenario against another. A failure to do this may lead to large amounts of disparate information entering the investigation and hindering the development of suspect sets and focused lines of enquiry.

Secondly, and arguably of more practical value, the combined package may help investigators prioritise offender groups. For example, where intelligence-led DNA screening is being considered, it might provide an effective mechanism for prioritising age groups to be sampled by, for instance, identifying the age bands where the offenders are most likely to be present. If we take the case presented in Figures 4.1 to 4.6 (a 30-39 white female manual worker with an unknown criminal record is stabbed in a rage/quarrel), the most likely age for the offender is between 30 and 39 years. The model also helps to order the second and third most likely age groups. This approach should, however, be used in conjunction with the frequency approach (which might point to a more narrow age range for the particular sub-group of victims), and other information thrown up by the investigation. It is important, therefore, to acknowledge that while the approaches used in Chapter 4 indicate some of the potential for the predictive approach, they also highlight some of the practical limitations of applying complex statistical approaches to real life predictive situations.

A final potential application could be as an interactive training tool. This might aid less experienced investigating officers to understand the complex picture of homicide, appreciating how knowledge of different variables may alter likely outcomes.

Additional thought would need to be given to further validating this analysis and then refining it in a way that might assist live investigations. It would, however, be possible for these predictive methods to be carried out in a relatively simply designed spreadsheet; alternatively, a web browser interface could be designed to provide an easy-to-use system for crime analysts. The victim characteristics would be entered on a web form, and the information sent to a remote computer. The information could be processed quickly, and bar charts and estimated profiles for the case would be returned, either automatically or via a crime analyst. Arguably the most appropriate place to locate this function would be within the NCPE Serious Crime Analysis Section (SCAS); their remit includes the provision of national investigative advice and support to hard-to-solve homicide investigations. A copy of the Homicide Index could reside within SCAS and be regularly updated, and the variables re-coded as described in this study. Furthermore, crime analysts within SCAS could be trained in interpreting statistical modelling process alongside the frequency approach at minimal cost.

This study has also highlighted a number of issues on the way in which homicide is described on the Homicide Index. These issues, in part, reflect the very complexity of describing acts of homicide. But it is also the case that the application of the Index to practical assistance to homicide investigations might benefit from a critical review of the way in which offences are coded. Many of these detailed coding points have already been considered by review of the Homicide Index and a revised data collection tool is currently being piloted (Mayhew, 2001).

Recommendations

This research has indicated that the Homicide Index has some potential in its application to assist on-going hard-to-solve (and unsolved ‘cold case’) homicide investigations. Consequently, the following points are recommended for consideration:

- A periodically updated copy of the Homicide Index should reside within the National Centre for Policing Excellence Serious Crime Analysis Section (SCAS) to complement the national analytical service provided to serious crime investigations.

- A user-friendly application of the statistical model should be developed and SCAS analysts should be trained in the application and interpretation of the models. Monitoring of the application and interpretation of findings would need to be undertaken as part of the process of validation.

- A template for a 'package' of investigative advice incorporating both the frequency and statistical modelling approaches should be developed by SCAS.

- A summary of the potential for this application should be included in updated advice on the running of DNA intelligence-led screens for homicide enquiries.

References

Aitken, C., Connolly, T., Gammerman, A., Zhang, G., Bailey, D., Gordon, R. and Oldfield, R. (1996) Statistical modelling in specific case analysis. *Science and Justice* 36 (4): 245-255.

Aitkin, M., Anderson, D., Francis, B. and Hinde, J.P. (1989) *Statistical Modelling in GLIM.* Oxford: Oxford University Press.

Aitkin, M. and Francis, B. (1992) Fitting the Multinomial model with continuous covariates in GLIM. *Computational Statistics and Data Analysis*, **14**, 89-97.

Brookman, F. and Maguire, M., (2003) Reducing Homicide: a review of the possibilities. On-line Report 01/03. London: Home Office.

Farrington, D.P. and Lambert, S. (2000) Statistical approaches to offender profiling. In D. Canter and L. Alison (Eds.) *Offender Profiling Series IV: Profiling property crimes.* Aldershot: Ashgate Publishing.

Feist, A. and Newiss, G. (unpublished) *Watching the detectives: analysing hard to solve homicide investigations.*

Flood-Page, C. and Taylor, J. (2003) *Crime in England and Wales 2001/2: Supplementary volume*. Home Office Statistical Bulletin 01/03. London: Home Office.

Francis, B. and Crosland, P. (2002) *The Police National Computer and the Offenders Index: Can they be combined for research purposes?* Home Office Online Report 04/02. London: Home Office.

Francis, B., Green, M. and Clarke, M. (1992) Model Fitting Applications in GLIM4. In: Advances in GLIM and Statistical Modelling. *Lecture Notes in Statistics,* 78. Springer Verlag, New York.

Francis, B., Green, M. and Payne, C. (1993) *The GLIM4 system release 4 manual.* Oxford: Oxford University Press.

Gibson, E. and Klein, S. (1961) *Murder.* London: H.M.S.O (Studies in the causes of delinquency and the treatment of offenders, 04).

Gibson, E. and Klein, S. (1969) *Murder, 1957 to 1968: a Home Office Statistical Division report on murder in England and Wales.* London : H.M.S.O, 1969. (Home Office research study 3).

Innes, M. (1999) The media as an investigative resource in murder enquiries. *British Journal of Criminology*, 39 (2): 269-286.

Karlsson, T. (1999) 'Findings on the victim of sharp-force homicide can predict inter-relationship with the perpetrator' *Forensic Science International*, 101 (1): 33-41.

Keppel, R.D. and Weis, J.G. (1993) 'HITS: Catching Criminals in the Northwest'. *FBI Law Enforcement Bulletin*, 62 (4) pp14-19.

Mayhew, P. (2001) *Review of Information on Homicide: A Discussion Document*. London: Home Office.

Soothill, K., Francis, B., Ackerley, E. and Collett, S. (1999) *Homicide in Britain.* Edinburgh: Scottish Executive.

Soothill, K., Francis, B., Ackerley, E. and Fligelstone, R. (2003) *Murder and serious sexual assault: What criminal histories can reveal about future serious offending* Police Research Series Paper 144 London: Home Office.

Wolfgang, M. (1966) *Patterns in Criminal Homicide*. New York: First Science Editions.

Appendix A: Definition of variables used in the study

The following variables used in the statistical analysis required re-coding into more manageable and useful numbers of categories. Below are the details of the necessary changes.

Job Status of victim (Homicide Index variables 'ECONPSTN' and 'KEYOCCPN')

New categories	**Original Homicide Index categories**		
	ECONPSTN	KEYOCCPN	
Manual		1	Prostitute
		2	Vagrant
		3	Police officer
		4	Prison officer
		5	Firefighter
		6	Ambulance staff, paramedics
		9	Security guards or other security staff
		12	Other manual occupation (i.e. skilled and semi-skilled manual, unskilled workers)
Non-manual		7	Social workers
		8	Medical staff (doctors, nurses etc.)
		11	Other non-manual occupations (i.e. higher or intermediate managerial or professional, and skilled non-manual)
Student	2		Student
Retired	4		Retired
Inactive	3		Unemployed
	5		Other adults economically inactive (not working or looking for work)
Other	6		Children under school age
	7		Not known
		10	Inmate of penal institution

This measure is a combination of the economic position and key occupation variables for the victim. For each case, only the value of one of the variables is used to determine the overall job status of the victim.

Relationship of victim to suspect (Homicide Index variable 'RELATION')

New categories	Original Homicide Index categories	
Offspring	1	Son, daughter (including adopted)
	2	Stepson, stepdaughter (including child of suspect's cohabitant/lover
Spouse/lover	4	Spouse
	5	Ex-spouse, estranged spouse
	6	Cohabitant, common-law spouse
	7	Ex-cohabitant, ex-common-law spouse
	8	Lover, mistress, sweetheart
	9	Ex-lover, ex-mistress, ex-sweetheart
	10	Lover's spouse, spouse's lover, cohabitant's spouse or lover, lover's lover
	11	Homosexual relationship – long-term
	12	Homosexual relationship – casual
Other family	3	Parent, step-parent
	13	Other family (including foster children)
Acquaintance	14	Criminal associate
	15	Friend, ex-friend
	16	Prostitute to client
	17	Commercial, professional or business relationship, where the *victim* was killed in the course of carrying out their occupation
	18	Commercial, professional or business relationship, where the *suspect* killed a client in the course of carrying out their occupation
	19	Other known (acquaintance)
Stranger	20	Police officer, prison officer killed in the course of their duty (takes highest priority)
	21	Stranger – Terrorist killing
	22	Stranger – Contract killing
	23	Stranger – Other
Unknown	24	Not known (insufficient information)

NB Code 25 – 'No current suspect' – is not included here since at least one suspect was found guilty for all cases in the sample.

Method of killing (Homicide Index variable 'METHOD')

New categories	**Original Homicide Index categories**	
Sharp instrument	1	Sharp instrument
Blunt instrument	2	Blunt instrument
Hitting/kicking	3	Kicking or hitting etc. without a weapon
Strangulation, asphyxiation or drowning	4	Strangulation
	12	Drowning
	16	Suffocation, asphyxiation or smothering
Fire	17	Arson – setting fire and causing death by fire
Shooting	7	Shooting
Other	5	Exhaust fumes (includes all carbon monoxide poisoning)
	6	Other poisoning (drugs etc.)
	8	Exposure of newly-born child (killed by natural elements)
	9	Negligence or neglect
	10	Aborting
	11	Explosion
	13	Causing to fall against a hard surface
	14	Burning, scalding
	15	Struck by motor vehicle
	18	Other (includes non-specific methods in baby battering cases i.e. shaking etc.)
Not known	19	Not known

Circumstances of homicide (Homicide Index variable 'CIRCMST')

New categories	Original Homicide Index categories
Rage or quarrel	1 Rage, quarrels, fights etc. involving related persons 2 Rage, quarrels, fights etc. involving non-related persons 4 Child abuse (neglect, excessive punishment, baby battering etc.)
Jealousy or revenge	3 Jealousy or revenge
Sexual	5 Sexual
Theft or other gain	6 Robbery 7 Burglary 8 Other gain
Feud	10 Faction fighting or feud (gangs or rival groups)
Imbalanced mind	20 Irrational act carried out by apparently insane or disturbed suspect
Other	9 Racial violence 11 Resisting or avoiding arrest, escaping from custody 12 Mercy killing, suicide pact 13 Prevent victim informing on suspect or testifying against suspect 14 Arson of property 15 Reckless act – motor vehicle 16 Reckless act – other 17 Homicide of mother arising from abortion or similar act 18 Terrorist incident (victim killed by terrorist activity) 19 Other circumstances/motive 21 Motiveless (sufficient information available to suggest that there is no rational motive)
Unknown	22 Not known (insufficient information)

Appendix B: Definition of offence types

In addition to the measure of whether a victim or offender had previous convictions for any type of offence, measures for three particular types of criminal career have been created: whether or not there are previous convictions for violent offences, sexual offences or drugs-related offences. These groupings have been made according to Section II of the Offenders Index codebook, using the Offenders Index variables denoting the main offence and the sub-classes of the offence, and are summarised here.

Sexual offences

Code	Offence
16	Buggery
17	Indecent assault on a male
19	Rape
20	Indecent assault on a female
21	Unlawful sexual intercourse with girl under 13
22	Unlawful sexual intercourse with girl under 16
23	Incest
25	Abduction
74	Gross indecency with a child
139	Indecent exposure
192	Gross indecency with children (1963-78)
18	Indecency between males
24	Procuration
26	Bigamy
27	Soliciting by a man (1978 onwards)
86	Possession of obscene material, etc. (1982 onwards)
107	Keeping a brothel
166	Offences by prostitutes
187	Living on prostitute's earnings etc. (1963-78)

Violent offences

Code	Offence
1	Murder
2	Attempted murder
3	Threat or conspiracy to murder
4	Manslaughter, etc.
920	Death or injury to person by dangerous driving (1.7.64-31.12.86)
921	Aiding, abetting, causing or permitting death or injury to person by dangerous driving (1.7.64-31.12.86)
37, sub 1	(1992 onwards) Aggravated vehicle taking
5	Wounding or other act endangering life
6	Endangering railway passenger
7	Endangering life at sea
8	Other wounding, etc.
9	Assault (until 1988)
10	Intimidation and molestation (until 1979)
11	Cruelty to or neglect of children
12	Abandoning children under two years
13	Child abduction
14	Procuring illegal abortion
15	Concealment of birth
35	Blackmail
36	Kidnapping
64	Rioting
65	Violent disorder
81	Firearms offences (1979 onwards)
103	Aggravated assault
104	Assault on a constable
105	Common assault (up to 1974)
109	Cruelty to or neglect of children

Drugs offences

Code	Offence
77	Misuse of Drugs (1972-92), Criminal Justice (International Co-operation Act 1990) (1993 onwards)
92	Misuse of drugs (1992 onwards)
93	Misuse of drugs (1993 onwards)
168	Offences in relation to Public Health
37	Possession of soft drugs (1968-1971)
193	Misuse of Drugs
195 subclass 5	Dangerous Drugs Acts 1965 and 1967 (up to 1971)

Appendix C: The Multinomial Logistic Model

The Multinomial Logistic Model is used for modelling outcome variables or characteristics which are categorical and with more than two categories. The model does not assume that the categories are ordered, so it is useful for modelling variables such as method of killing.

The method models the chance or probability that an offender will have a particular category of the variable being modelled. This probability will depend on a set of victim variables. If the variable being modelled has five categories, then there are five probabilities – however the probabilities sum to one, and so we only need to estimate four probabilities from the model.

We now define the model mathematically. We assume that the outcome variable for case *i* has *R* categories with associated probabilities p_{ir}, *r=1,2,…,R*. In other words, for every case, the probabilities $\sum_{r=1}^{R} p_{ir} = 1$ category sum to one.

This constraint means that there are only *r-1* distinct probabilities to estimate.

We model the probabilities by constructing a set of multinomial logits for each case – transformation of the probabilities.

$$\theta_{ir} = \log(p_{ir} / p_{i1}) \qquad \text{r=1,…R}$$

with the first category being used as the reference category. The above implies that $\theta_1 = 0$, so there are only (r-1) distinct logit parameters. The advantage of using the logit transformation is that the θ_{ir} are allowed to take any value – positive or negative.

For each θ_{ir} we build a separate statistical model, relating the logit for the *r*th category to a set of explanatory victim variables As these victim variables are categorical, then we construct a set of (0,1) dummy variables to represent the categories for each variable. If, after this expansion, there are *P* explanatory dummy variables X_{ip} *p=1…P* , then the statistical model is given by

$$\theta_{ir} = \beta_{0r} + \sum_{p=1}^{P} \beta_{pr} X_{ip} \qquad r = 2...R$$

In order to get estimates of these parameters in GLIM, we need to use the relation between the multinomial and Poisson distributions. The procedure for doing this is described in Aitkin *et al.* (1989), Francis and Green (1992) and Aitkin and Francis (1992).

Produced by the Research Development and Statistics Directorate, Home Office

This document is available only in Adobe Portable Document Format (**PDF**) through the RDS website

Home Office
Research, Development and Statistics Directorate
Communication Development Unit
Room 264
50 Queen Anne's Gate
London SW1H 9AT

Tel: 020 7273 2084 (answerphone outside of office hours)
Fax: 020 7222 0211
Email: publications.rds@homeoffice.gsi.gov.uk

ISBN 1 84473 282 7