

## Statistical Modelling in the Investigation of Stranger Rape

MIRJAM TER BEEK<sup>1,\*</sup>, PAUL VAN DEN ESHOF<sup>2</sup> and BAS MALI<sup>3</sup>

<sup>1</sup>*Criminal Investigation Department, Police Region Utrecht, the Netherlands*

<sup>2</sup>*National Police Agency, Zoetermeer, the Netherlands*

<sup>3</sup>*Police Academy, Apeldoorn, the Netherlands*

### Abstract

*A sample of stranger rape offences (n = 271) registered in the Dutch Violence Crime Linkage Analysis System database in the Netherlands between 1997 and 2007 was studied with the objective of developing statistical models, which give an indication of the probability of basic offender characteristics. Observable crime characteristics concerning the modus operandi, interaction between the offender and the victim, violence, precautionary measures, and sexual behaviours were selected in the dataset. Offender characteristics were selected based on their usefulness for the police organisation in narrowing the scope of a criminal investigation. Spatial behaviour, criminal history, and living situation of the offender were selected. From the predictive models, four out of five achieved a correct rate of over 70%, and all models predicted better than the best guess method. The proposed models for distance and prior convictions for violence seem particularly promising. Both these models show an improvement of correctly predicted offender characteristics of more than 20 percentile points compared with that which could have been estimated based on the average in the total sample. The predictive value of the models needs to be tested further with 'new offences', which were not used to construct the model. In general, the current study supports the finding that crime characteristics can be used to get an indication of the probability of certain offender characteristics. Nevertheless, for an understanding of the relationship between the crime characteristics and offender characteristics, a further development of a theoretical framework is urgently necessary. Copyright © 2009 John Wiley & Sons, Ltd.*

**Key words:** stranger rape, statistical modelling, spatial behaviour, criminal history, living situation

\*Correspondence to: Mirjam ter Beek, Police Region Utrecht, Criminal Investigation Department, Briljantlaan 3D, 3523 CD Utrecht, the Netherlands.

E-mail: mirjam.ter.beek@utrecht.politie.nl

## INTRODUCTION

Experienced detectives know that in investigations of serious crimes, an effective gathering of information during the first 24/48 hours is of crucial importance. Nevertheless, in practice, the investigation during these first hours often lacks a systematic approach. Decisions are needed on how to focus the investigation. In the case of a stranger rape offence, the information upon which to decide the direction of the investigation is often limited. There is no relationship or social connection between the offender and the victim, and in most cases, there are few or no witnesses. In addition, sexual offenders constitute a markedly heterogeneous group (Knight & Prentky, 1990; Righthand & Welch, 2001; Wijk, Mali, Bullens, Prins, & Klerks, 2006). What kind of suspect are we looking for? What was the motivation of the suspect? In these situations, it is difficult to direct the investigation, as the potential suspect can basically be anyone.

### Classification of rapists

To be able to narrow the group, one possibility is to distinguish between different types of sexual offenders. Different researchers have tried to develop a method to accurately distinguish between different types of sexual offenders (Groth, Burgess, & Holstrom, 1977; Knight, 1999; Knight & Prentky, 1990). Such a distinction could be useful in the treatment of offenders. These kinds of classifications give us a theoretical framework to help us understand stranger rape offenders. Unfortunately, they do not necessarily give practical directions as to how to direct the investigation. One of the difficulties with most of these studies is that the independent variables used to classify the offender do not solely consist of information, which is known to the police at the beginning of an investigation. Another frequently applied method is thematic behavioural analysis (Canter, Benell, Alison, & Reddy 2003; Häkkinen, Lindlöf, & Santtila, 2004; Kocsis, 2006). Rapists are classified based on overt crime scene behaviour, which is grouped into different behavioural themes. However, only a few of these studies have been able to find a correlation between the behavioural themes and offender characteristics (House, 1997; Kocsis, 2006; Santtila, Junkkila, & Kenneth-Sandnabba, 2005). Therefore, the applicability of this method for investigations is also limited as long as a link with behavioural characteristics is not forthcoming. Another criticism of behavioural analyses is that combining different crime scene variables and offender characteristics is not specific enough for use in criminal investigations.

### Pragmatic approaches—prioritisation techniques

In practice, in the Netherlands, there is an emphasis on analysing and interpreting of separate crime scene variables (method of approach, verbal behaviour, sexual behaviour, use of violence), rather than classifying the offender as a certain rapist type to whom a standard set of characteristics are subsequently attributed (Eshof & Nierop, 2007). Currently, we are again seeing research in the field of rape, where the predictive value of the separate crime characteristics for the probability of specific offender characteristics is studied (Wilson & Alison, 2005). In this same study, Wilson and Alison (2005) have shown that single crime scene variables have been able to significantly alter the probability of specific offender characteristics that can be used in police investigations of stranger rape. Also, the Department of International Police Information of the National Police

Agency has found a number of 'knowledge rules', relationships between single crime scene variables and offender characteristics, which can be applied in investigations (Eshof & Nierop, 2007; Jackson, Eshof, & Kleuver, 1997). Although the method does not provide a classification model for different types of rapists, it does provide a pragmatic method to be able to prioritise lines of enquiry in an investigation.

The pragmatic approach of looking for relationships between single crime scene variables and offender characteristics corresponds with the study of Aitken to prioritise suspects of sexually motivated child murders (Aitken, Connolly, Gammernan, Zhang, Bailey, Gordon, & Oldfield, 1996). However, Aitken has taken this type of research one step further with the development of statistical models using a multivariate approach for the prediction of offender characteristics in sexually motivated child homicides. For his research, Aitken used the CATCHEM database, which consists of victim and crime scene variables available in police investigations. The first set of models Aitken developed were designed to predict only the most basic of offender characteristics, respectively the offender's age (0–21 or over 21), previous conviction status (yes or no), the proximity of the crime to offender's residence (within 5 miles or further away), marital status (living with a partner or not), and whether the offender was known to the victim (yes or no). Aitken used a computerised statistical procedure to develop the best model for each offender characteristic. He seems to have developed a method that can provide investigators with probabilistic statements about characteristics of the offender based on crime scene variables available to the police. The model is meaningful in prioritising lines of enquiry in an investigation. Although the study of Aitken is based on the analysis of sexually motivated child homicides, the findings are applicable to other types of crimes, provided there are appropriate data sets of solved cases (Aitken, Connolly, Gammernan, Zhang, & Oldfield, 1995). Karlsson (1999) and Francis, Barry, Bowater, Miller, Soothill, and Ackerley (2004) further showed the benefits of a statistical modelling approach by predicting the probabilities of offender characteristics in homicide cases. Safarik, Jarvis, and Nussbaum (2002) and Safarik and Jarvis (2005) developed such models for sexual homicide of elderly females. A similar empirically validated model to aid law enforcement in prioritising lines of enquiry for rape cases committed by stranger offenders will be of great value to criminal investigations.

If we are able to predict the probable nature of some of the basic offender characteristics, this can provide the investigators with significant and useful information to focus the search for a stranger sexual offender. It will not exclude suspects with certain characteristics. It tells us the likelihood of certain offender characteristics and therefore will help the investigators to prioritise.

One of the first applications of a statistical model for suspect prioritisation for stranger rape was developed by the Criminal Behaviour Analysis Unit (CBAU) in Newfoundland, Canada. They developed a computerised method for suspect prioritisation for a range of crime types, the Criminal Suspect Prioritisation System (CSPS) (House, 1997). One of the prioritisation systems was developed with a sample of stranger rape crimes ( $n = 60$ ). As with the statistical models of Aitken *et al.* (1995), this prioritisation system was also developed to contribute in a meaningful way to the investigative process. The researchers used police criminal records, crime scene information, and victim/witness statements as a primary source of data to develop their prioritisation system. A review of international literature led them to conclude that criminal antecedents, age range of the offender, and spatial actions of the criminal were characteristics with potential for the prioritisation of offenders (House, 1997). They used smallest space analysis to determine the behavioural

themes in their sample. The CSPS is an established system within the CBAU to assist in the investigation of crimes that range from burglary and robbery to rape and murder (House, 1997). Unfortunately, the model has not been empirically tested. It is not clear if crime behaviours improved the prediction, or if an estimation of the offender characteristic based on the most common characteristic of the offenders in the dataset would have achieved similar results ('best guess method').

Another application of a statistical prioritisation instrument in stranger rape offences was developed by Davies, Wittebrood, and Jackson (1998). They were able to develop a statistical model for the probability of an offender having previous burglary convictions, violence convictions, and the probability of an offender being a one-off sexual offender. Unfortunately, the improvements in the prediction of the models in relation to what could have been estimated based on the average of the group were modest. The best results were shown for the model predicting burglary conviction. This model showed an improvement of 0.13 percentile points compared with the best guess method.

### Practical value

As mentioned earlier, several initiatives in suspect prioritisation techniques based on crime characteristics (Aitken *et al.*, 1995; House, 1997; Karlsson, 1999; Safarik and Jarvis, 2005; Safarik *et al.*, 2002), best guess method (Snook, Wright, House, & Alison, 2006), or a combination of both (Francis *et al.*, 2004) have been developed and successfully used in practice. Nevertheless, the development of a prioritisation technique for practical application in stranger rape investigations is still premature. Is there an added value in such a method? Detectives have to make quick decisions. They are forced to narrow their investigation based on time and capacity restrictions. Therefore, they have to work with presuppositions. However, intuitive knowledge based on practical experience is proving to be insufficient as a guideline for police decision making (Bayley & Bittner, 1984 cited in Yarmey, 1990). Domain-specific knowledge based on probabilities is necessary. Detectives seem to lack the necessary skills to infer characteristics of the offender (Adhami & Browne, 1996; Jackson *et al.* 1997; Kocsis, Hayes, & Irwin, 2002; Kocsis, Irwin, Hayes, & Nunn, 2000). In addition, to infer characteristics without empirical support is currently being criticised (Snook, Cullen, Benell, Taylor, & Gendreau, 2008). The statistical model is a tool that can assist in predicting the probabilities of basic offender characteristics. The model will not be able to give certainties, but the expectations that are given are empirically researched and not based on prejudices. The assumptions are therefore more realistic and reasonable in terms of their chances of being correct. In other words, the model can support detectives to base their presuppositions on expectations that are *most likely* to be true. With the aid of the current model, they can base their expectations upon the total sample in this study of 271 stranger rape cases rather than just the few cases they have dealt with in their own career. The predictions can be used for setting suspect parameters, prioritising searches, or developing lines of enquiry at the start of an investigation. Nevertheless, there is always the risk that the direction of the investigation based on the possible offender characteristics turns out to be wrong, and the offender needs to be found by going in a different direction. We still work with probabilities and have to keep in mind that the offender might not match the most likely characteristic. The model will therefore not exclude suspects with certain characteristics. It will give us the likelihood of certain offender characteristics and will help the investigative team with the screening of offender subgroups. Overall, claims should be examined critically. One of the possibilities is to use

Toulmin's (1958) philosophy of argument as a framework to evaluate the claim being made (Alison, Smith, Eastman, & Rainbow, 2003). One of the constituents of this framework is whether there is 'rebuttal': is there a condition during the crime under which the claim may not be true. Not every aspect of a crime can be captured by empirically researched knowledge rules or be included in the statistical model. Therefore, it is important to take into account the entire offence and see if certain conditions appear under which the claim may not be true. van den Eshof and Nierop (2007) find it unwise to base decision making exclusively on knowledge rules. They argue that for a complete picture, it is necessary to reach conclusion based on the interpretation of the other characteristics of the offence. The added value of the statistical model is that it can provide the detective with a direction on how to focus the investigation at the very start. It can therefore make the first days of an investigation more efficient. The model should *not* be used in isolation but in conjunction with other advice/evidence provided by those involved with the investigation; the model is there to *assist*. For a more tailored behavioural advice, a behavioural investigative advisor should be consulted.

The overall aim of the study was to develop statistical models for the prediction of basic offender characteristics in offences of stranger rape. The models need to be useful in the prioritisation of lines of enquiry in police investigations. In other words, the offender characteristics needed to be useful in narrowing the scope of a criminal investigation. Also, the models should consist of a reasonable number of variables to be workable in investigations. An additional aim was to see which models were able to predict the probability of the offender characteristics with greater accuracy as compared with what could be achieved by guessing based on the average of the group. Therefore, insight into the average of the offender characteristics in the total sample was also necessary. The final aim was to see which crime scene characteristics were included in the models and what their contribution was in predicting the probability of the offender characteristic.

## METHOD

### Cases

In this study, 271 cases of stranger rape offences registered from 1997 to July 2007 in the Dutch Violence Crime Linkage Analysis System (ViCLAS) were used. ViCLAS is a Canadian database program that was developed to serve as a resource for the investigation of serial sexual offenders. It is used as an operational instrument. Information regarding sexual offences, the victim, and the behaviour of the offender are stored in ViCLAS. The data is mainly based on the victim's statement, and therefore contains information that is generally available to detectives in an investigation.

The rape offences were selected based on the following criteria: the victim was a female, was older than 16 years, survived the offence, did not have a relationship or other social connection with the offender, and the offender was found guilty under Article 242 of the Dutch Criminal Code of Conduct (sexual penetration of the body through violence, threat, of violence or another fact). Cases in which the offender only kissed the victim were not selected. Although in legal terms, so-called French kissing constitutes rape, the fact that the offence is less serious and there is most likely only limited information about the behaviour of the offender during the crime led to the decision not to select these offences. Offences with multiple offenders were excluded as well. If an offender was registered with more than one offence in the database, the so-called serial offenders, the latest offence in

the series was included. Latest offences were selected in line with Davies' (1992) suggestion that the later offences in a series could well be more revealing, using more precautionary measures than earlier ones, and therefore less affected by situational factors. In this study, in 46% of the cases, the rape was part of a series. Each series consisted on average of three rapes. The serial offenders did not significantly differ from the 'one-off' rapists in the selected offender characteristics. It is likely that some 'one-off' rapists may well be serial offenders whose other offences were either not reported to the police or did not lead to a conviction. The selected 271 rape cases represent the total sample registered in the ViCLAS database from 1997 to July 2007 meeting the above-mentioned criteria. Nevertheless, some stranger rape cases will most likely be missed. Although the aim is for all serious sexual offences in the Netherlands to be registered in the national ViCLAS database, this depends upon the input from police regions. In some instances, the sexual offence is investigated in the region, but the National Police Agency is not informed, and therefore registration in the national ViCLAS database does not take place.

### Crime characteristics

In this research, the emphasis was on applicability for the police organisation; therefore, we used crime scene information and victim statements as a primary source of data from which to develop what will be termed 'crime characteristics'. Crime characteristics were selected from the ViCLAS questionnaire based on the following criteria: the victim was able to provide the police information about the crime characteristic, *or* the police could infer the crime characteristic from the crime scene. In addition, the crime scene behaviour was found to have or likely to have predictive value in relation to offender characteristic(s).

Information concerning the modus operandi, offender–victim interaction, violence, precautionary behaviours, sexual behaviours, and a few offender characteristics that could be observed by the victim were included in the data set. To facilitate data analysis and interpretation, for the purposes of the current study, some variables with multiple answer categories were subdivided, or conceptually similar categories for each variable were merged into groupings of two. Categories were constructed with a view to produce dichotomous variables. Most variables were recoded on a presence–absence basis, whereas others were recoded into less–more type categories. In general, in the ViCLAS database, only the *presence* of crime characteristics is registered. In this study, if a variable was not registered as present, it was coded as 'not present'. Therefore, the category 'not present' is broader than offences where the crime characteristic was not present, but also contains those offences in which the victim did not observe the crime characteristic, or the specific question was not asked or recorded by the police officer. Variables that were not able to discriminate were deleted from the database. A variable was considered not be able to discriminate if less than 5% (or more than 95%) of the offenders behaved in this particular way. As a result of the variable condensation process, some 168 variables were reduced to a final set of 71 variables. For a complete overview of the variables, the reader can contact the first author of the paper.

### Offender characteristics

For the development of the statistical model, offender characteristics were selected based on their usefulness for the police organisation in narrowing the scope of a criminal investigation.

Their usefulness was applied to the extent that the offender characteristics needed to be known to the police. In other words, the police needed to have access to the registration of the offender characteristics in general to be able to use the characteristic in narrowing the scope of the investigation. In addition, the offender characteristics needed to discriminate between offenders. If a certain characteristic is present in almost every stranger rape offender, then it will not be very useful in narrowing the search. In relation to the statistical model, a substantial number of predictions would already be predicted correctly using a strategy to predict the dominant characteristic for all offenders. This would make it difficult to develop a model that could improve the prediction rate, which could be achieved based on the average of the total group. Another aspect of the application of the offender characteristic's usefulness was that crime scene behaviour needed to be predictive of the probability of the presence of the offender characteristic. A final criterion was that the victim was not able to provide the police with the information concerning the offender characteristic. For stranger rape cases, the victim is usually able to provide information concerning the offender, such as their relationship with them, an indication of their physical appearance, whether more than one offender was involved, and an estimation of the offender's age. Only two offender characteristics available in the ViCLAS database do met all four criteria: spatial behaviour of the offender and his living situation.

### **Spatial behaviour**

Details on an offender's residence are routinely part of police records, which makes these findings potentially useful. In addition, the discriminative power can be guaranteed, as the distance travelled can be chosen in such a way that offenders who meet the victim within the chosen distance and those who travel further can be equally distributed. The estimation of whether the offender's residence is in proximity to the crime location or meeting point with the victim can effectively narrow an investigation. Geographic profiling systems have already shown to be useful for crime investigations. However, those systems are most frequently used for serial offences, as a certain number of crime locations are necessary in order to predict the likelihood of an offender's residence (Rossmo, 2000, cited in Snook *et al.*, 2006). If we want to predict the likelihood of the proximity of an offender's residence with a single offence location, other parameters are needed. Other research of stranger rape offences has shown that crime characteristics do influence the likelihood that the offender lives close to the crime location (Kleuver *et al.*, 1996; Santtila, Laukkanen, & Zappalà, 2007; Santtila *et al.*, 2005; Wilson & Alison, 2005).

### **Living situation**

The living situation of the offender is another offender characteristic that is interesting when applied to directing a criminal investigation. In this study, living situation refers to whether the offender is cohabiting with a partner or not. The police have access to the registration of occupants' home addresses, although we have to keep in mind that conclusions are dependent on a correct registration of the address of the offender in any one local authority's database. A study by van den Eshof and de Kleuver (2000) showed that cohabitation with a partner is reasonably equally divided. They found that in a sample of stranger rapists, 43% were co-habiting with their wife or girlfriend (cited in Eshof and Nierop, 2007). Living situation has rarely been studied in relation to crime scene behaviour. Nevertheless, Kocsis (2006) did find a relationship between different criminal behaviour patterns and the likelihood that the offender was in some form of conjugal relationship.

A further exploration of possible correlations between crime scene characteristics and the offender's living situation is of great value for criminal investigations.

### Previous convictions

A previous conviction of the offender is another offender characteristic that has potential for suspect prioritisation, as described by Snook *et al.* (2006). Convictions are routinely registered in judicial databases, which are available to the police. Previous convictions in general will not have enough discriminative power, as studies of sexual offenders have shown that the majority have a previous criminal history (Eshof, Kleuver, Ho Tham, & Zwiers, 1995, cited in Frenken, Gijs, & Beek, 1996; Sothill, Francis, Ackerley, & Fligelstone, 2002). Therefore, a simple distinction between the presence and absence of a criminal history is not very useful for suspect prioritisation. Consequently, we need to look more specifically at the type of previous convictions. Different researchers have found a relationship between crime characteristics and previous criminal convictions of stranger rapists (Davies *et al.*, 1998; House, 1997; Wilson & Alison, 2005). Further insight into the relationship between behaviours exhibited by the offender during a stranger rape and previous criminal convictions of these offenders will provide us with useful knowledge for prioritising the direction of the investigation. Therefore, models for the probability of previous convictions for sexual, violence, or property crimes are being developed.

### Analytical process

The presence of the selected offender characteristics in the total sample was studied. The linear distance was used to measure the distance between the home address of the offender and the site of the first encounter with the victim. Cases for which the home address of the offender at the time of the offence was unknown ( $n = 18$ ) were left out of the analysis. For the other 253 cases, the distance was calculated by using geographical coordinates of the postal codes of the home address of the offender and the address of the first encounter. The living situation of the offender at the time of the offence was known for 198 offenders, and those cases were used to develop the model that predicts the probability of the offender cohabiting with a partner. Court information was studied to determine whether the offenders had previous convictions. Pre-conviction details for all offenders ( $n = 271$ ) were available. No distinction was made between a conviction for an attempted or a completed offence. Both were registered as a conviction. For a complete overview of the different types of convictions, the reader can contact the first author of the paper.

In order to predict the probability of the offender characteristics with multiple independent variables, we used the logistic regression model. A logistic regression model calculates the probability of one of the two categories of a dichotomous variable (offender characteristic) given independent variables (crime characteristics). For a further explanation of the statistical procedure logistic regression, the reader is referred to statistical manuals.

## RESULTS

### Distance equation

A statistical model 'distance' was constructed to predict the probability that the offender lives within a 3-km radius of the site where the first encounter with the victim took place.

In accordance with earlier empirical results, most of the stranger rape offenders lived in close proximity to the initial point of contact with the victim. Fifty per cent of the stranger rape offenders lived within 3 km of the first encounter site, and 74% within 10 km of this site.

In the logistic regression equation, the crime scene characteristics 'the victim is a prostitute' and 'the crime scene is a public place' strongly and negatively contributed to the likelihood of the offender living within 3 km of the first encounter site. The finding that the type of location of the first encounter site influences the probability that the offender lives in close proximity to that site has also been found in previous research (Kleuver, Eshof, & Zweirs, 1996). In the current study, a public place was defined as an area freely accessible to the public, including stations for public transport, as well as public buildings, such as schools or hospitals. In keeping with the earlier findings of Wilson and Alison (2005), the con approach also reduced the chance that the offender lived within 3 km and was included in the equation. Other crime characteristics that negatively contributed and were included in the equation were: 'the offender used a vehicle', 'the crime scene was the residence of the victim', 'the offender appeared to be angry', and 'the offender reassured the victim'. By contrast, if the crime scene was a residential area, if the offender used more force than necessary and/or if the offender negotiated with the victim, the probability that the offender lived within 3 kilometres of the initial contact site increased (Table 1).

Table 1. Variables in the distance equation

| Offender lives within 3 km of the first encounter |        |
|---------------------------------------------------|--------|
| Constant                                          | 0.267  |
| Victim is a prostitute                            | -1.314 |
| Offender used a vehicle                           | -0.961 |
| Crime scene residence victim                      | -0.855 |
| Crime scene public place                          | -1.465 |
| Scene residential area                            | -1.000 |
| Con approach                                      | -0.572 |
| Force more than necessary                         | 0.978  |
| Anger                                             | -0.909 |
| Reassured victim                                  | -1.120 |
| Negotiation with victim                           | 1.066  |

The statistical model 'distance' achieved an overall correct prediction correct of 70.8% with 10 variables in the equation. Offenders who were not living within 3 km were predicted correctly in 66.9% of cases, whereas offenders who were living within 3 km were correctly predicted in 74.6% of cases.

We will illustrate the application of the statistical model 'distance' with the following two cases.

An offender raped a female in a residential area and used more force than necessary to control the victim. The other variables in the distance equation were not present or unknown at the time of the investigation. The available information was entered into the equation, and we calculated that the probability that the offender lived within 3 km of the first encounter site was 90%. If further information about variables in the equation becomes available at a later date, this can be entered in the equation to calculate the most up-to-date probability that the offender lives within 3 km of the first encounter site.

In another case, the offender used the con approach to make contact with the victim, and during the offence tried to reassure the victim. (The other variables in the equation were again not present or unknown.) In this scenario, the probability that the offender lived within 3 km of the first encounter site was 19%.

### Equation living with partner

In the current study, the living situation of stranger rape offenders was known for 198 offenders. Of these offenders, 26% were living alone, 28% were living with parents or caretakers, 28% were cohabiting with a partner (and their children), and 18% were living together with someone other than the above-mentioned groups. A statistical model 'living with partner' was constructed to predict the probability that the offender lives with a partner.

In the logistic regression equation, we saw that if the offender was younger than 21 years of age, this negatively contributed to the probability that the offender lived with a partner. In contrast, if the offender complimented the victim, the probability that the offender lived with a partner increased. This seems to contradict the finding of Kocsis (2006) who found a relationship between his brutality pattern and the likelihood that the offender was in some form of conjugal relationship. The brutality pattern of Kocsis (2006) consisted of crime scene behaviours that had a very violent and forceful nature (Table 2).

Table 2. Variables in the equation living with partner

| Living with partner | B      |
|---------------------|--------|
| Constant            | -0.668 |
| Complimented victim | 1.183  |
| Age offender <21    | -2.062 |

The statistical model 'living with partner' achieved an overall correct prediction of 75.5%, with two variables in the equation. Offenders who did not cohabit with a partner were correctly predicted in 97.7% of cases, whereas offenders who did cohabit with a partner were correctly predicted in 13.5% of cases.

### Previous convictions of rapists: equation violence convictions

An analysis of the different types of prior convictions in the current study of stranger rapists ( $n = 271$ ) showed that 76% of offenders had been previously convicted, and 51% of the offenders had previously been convicted of a violent offence. These two findings are generally in agreement with a voluminous study ( $n = 1057$ ) of criminal careers of rape offenders in England (Soothill *et al.*, 2002).

A statistical model 'violence convictions' was constructed to predict the probability that the offender had previous convictions for violence. In the logistic regression equation, we saw that if the offender was younger than 21 years of age, if the victim was walking or cycling, if the crime scene was indoors, if the offender used a vehicle, if the offender took precautions to hide his identity, and/or if DNA was found at the crime scene, there was a negative contribution to the likelihood that the offender had previous convictions for

violence. In contrast, if the offender stole an item from the victim, if he was intoxicated, if he strangled and/or demeaned the victim, if he forced the victim to disrobe herself, and/or if he performed cunnilingus on the victim, the probability that the offender had prior convictions for violence increased (Table 3).

Table 3. Variables in the equation violent convictions

| Violence convictions                     | B      |
|------------------------------------------|--------|
| Constant                                 | 0.159  |
| Theft during rape                        | 0.933  |
| Victim's activity was walking or cycling | -0.699 |
| Offender intoxicated                     | 1.396  |
| Used a vehicle                           | -1.029 |
| Crime scene indoors                      | -0.641 |
| Strangled                                | 0.840  |
| Demeaned                                 | 0.736  |
| Victim disrobed herself                  | 0.669  |
| Precautions to hide identity             | -0.910 |
| DNA found                                | -0.969 |
| Cunnilingus                              | 0.900  |
| Offender <21 years                       | -0.799 |

The statistical model 'violence convictions' achieved an overall correct prediction of 71.7 % with 11 variables in the equation. Offenders who did not have any convictions for violence were predicted correctly in 67.9 % of cases, whereas offenders who did have convictions for violence were correctly predicted in 75.4% of cases.

We will use a case example to further explain the application of the model. A victim was attacked when she was walking home at night. The offender was obviously intoxicated. He strangled her and ripped her clothes. The offender raped her vaginally. During the entire assault, he was verbally very demeaning towards the victim. He stole her purse before he left. In this scenario, the probability that the offender has prior convictions for violence is 97%.

### Previous convictions of rapists: equation rape convictions

Only 29% of offenders had previous sex crime convictions and in 26% of cases previous rape convictions. The percentage of offenders that had previous sex crime convictions differs from earlier findings in which 42% of stranger rapists had previously committed sex crimes (Eshof *et al.*, 1995 cited in Frenken *et al.*, 1999). However, in the current study, a distinction was made between previously committed sex crimes and a previous conviction for a sex offence. The distinction between a previously committed sex crime and a previous conviction was made because the aim of this research is to develop a model with a practical application for the police. It is important to be able to distinguish between scenarios whether we are looking for a suspect with previous sex crime convictions or whether the offender we are looking for may not have previous sex offences in his criminal record, but might be responsible for a series of unsolved sex offences. The probability that the offender we are looking for has committed a previous sex offence (60%) is twice as high as the percentage of offenders with a previous conviction for a sex offence (29%).

A statistical model 'rape convictions' was constructed to predict the probability that the offender had earlier rape convictions. In the logistic regression equation, we saw that if the offender negotiated with the victim, if he communicated about himself, if he took precautions towards the victim, if DNA was found at the crime scene, if he displayed 'gentle' sexual behaviours such as fondling, kissing, and hugging, and/or if he exhibited a sexual dysfunction, then there was a negative contribution to the probability that the offender had a previous rape conviction. In contrast, if the offender was intoxicated, if he beat the victim, if he asked questions about the victim's personal life, if he forced the victim to disrobe herself, if he digitally penetrated the victim, and/or if the offender was white, the probability that the offender had prior rape convictions increased (Table 4).

Table 4. Variables in the equation rape convictions

| Rape convictions                       | B      |
|----------------------------------------|--------|
| Constant                               | -1.610 |
| Offender intoxicated                   | 0.711  |
| Beat victim                            | 0.746  |
| Negotiation with victim                | -0.449 |
| Communicated about himself             | -0.927 |
| Asked questions victim's personal life | 1.000  |
| Victim disrobed self                   | 0.948  |
| Precautions towards victim             | -0.846 |
| DNA found                              | -0.915 |
| 'Tender' sexual behaviours             | -0.740 |
| Digital penetration                    | 0.607  |
| Sexual dysfunction                     | -1.072 |
| Offender is white                      | 0.862  |

The statistical model 'rape convictions' achieved an overall correct prediction of 77.9% with 12 variables in the equation. Offenders who did not have any rape convictions were predicted correctly in 96% of cases, whereas offenders who did have convictions for sex crimes were correctly predicted in 27.1 % of cases.

### Previous conviction of rapists: equation property convictions

In the current study, 54% of offenders had previous convictions for property crimes.

A statistical model 'property convictions' was constructed to predict the probability that the offender had prior 'property convictions'. In the logistic regression equation, we saw that if the offender used a vehicle, if the crime scene was indoors, if the offender did not respond violently to the physical resistance of the victim, if DNA was found, if anal rape occurred, and/or if the victim estimated the offender to be younger than 21 years of age, there was a negative contribution to the chance that the offender had a previous conviction for property crimes. In contrast, and in accordance with the findings of Davies *et al.* (1998), if the offender stole items, if he committed the rape during a burglary, and/or if he was intoxicated, the probability that the offender had previous convictions for property crimes increased. If the offender used a 'blitz' approach, if he strangled the victim, and/or if he communicated about himself, the probability that the offender had previous convictions for property crimes also increased (Table 5).

Table 5. Variables in the equation property convictions

| Property convictions                       | B      |
|--------------------------------------------|--------|
| Constant                                   | 0.289  |
| Stool items                                | 0.654  |
| Rape during burglary                       | 1.404  |
| Offender intoxicated                       | 1.316  |
| Used vehicle                               | -0.916 |
| Crime scene indoors                        | -0.757 |
| Blitz approach                             | 1.030  |
| No violent response on physical resistance | -0.812 |
| Strangled                                  | 0.621  |
| Communicated about himself                 | 0.597  |
| DNA found                                  | -0.879 |
| Anal rape                                  | -0.722 |
| Age offender <21                           | -0.684 |

The statistical model 'property convictions' achieved an overall correct prediction of 67.7% with 12 variables in the equation. Offenders who did not have any prior convictions for property crimes were predicted correctly in 58.5% of cases, whereas offenders who did have prior convictions for property crimes were predicted correctly in 75.3% of cases.

## DISCUSSION

Mokros and Alison (2002) were unable to find support for the underlying assumption of offender profiling, the homology assumption. The homology assumption can be described as the phenomenon whereby if two persons commit a particular type of crime in roughly the same way, then they should be fairly similar to each other. If the homology assumption needs to be rejected, then any method, including statistical modelling, in which offender characteristics are predicted with crime scene variables, is debatable. However, Mokros and Alison (2002) tested the validity of the homology assumption with respect to a variety of arbitrarily chosen crime characteristics and offender characteristic simultaneously, therefore ignoring the fact that the homology assumption could be valid for a specific combination of crime scene characteristics and a single-offender characteristic. Obviously, not all socio-demographic features of the offender can be derived from the crime scene behaviour, and especially not every arbitrarily chosen combination of those offender characteristics. As mentioned before, in practice, in the Netherlands, there is an emphasis on analysing and interpreting separate crime scene variables (method of approach, verbal behaviour, sexual behaviour, use of violence), instead of classifying the offender as a certain rapist type to whom a standard set of characteristics are subsequently attributed.

The current study confirms previous findings that crime characteristics can be used to predict the probability of a single offender characteristic in stranger rape offences (Davies, *et al.*, 1998; House, 1997; Kocsis, 2006; Kleuver *et al.*, 1996; Santtila *et al.*, 2005, 2007; Wilson and Alison, 2005). We succeeded developing of statistical models for the prediction of the probability of offender characteristics of stranger rapists meeting previously set criteria. From the predictive models, four out of five achieved a correct prediction rate

of over 70%, and all models did better than the 'best guess method'. The proposed models for 'distance' and 'violence convictions' seem particularly promising. Both these models show an improvement of correctly predicted offender characteristics of more than 20 percentile points compared with that which could have been estimated based on the average in the total sample. The practical value of the models 'rape convictions' and 'living with partner' is limited, as the correct prediction is slightly better than that which could have been guessed. In other words, since in the total population only 26% of the offenders are convicted for a previous serious sexual assault, predicting that none of them will have a prior conviction for a serious sexual assault will already be correct in 74% of the cases. Therefore, the correct prediction rate of 78.2% for rape convictions is only slightly better than what could have been achieved by guessing. Unfortunately, the statistical models do not provide a great deal more insight into our understanding of the relationship between the crime characteristics and the offender characteristics. In the statistical model, the crime characteristics together determine the probability of the offender characteristic. The single crime characteristics can have a minor influence when studied on its own, whilst in the statistical model, together with the other crime characteristics, it was shown to contribute significantly. In general, already existing theoretical frameworks for the relationship between crime behaviours and offender characteristics are insufficient to interpret this relationship between the crime characteristics and the offender characteristics. Studies that are able to identify theoretical constructs for understanding the links between the crime characteristics and offender characteristics are urgently needed.

Nevertheless, the availability of the statistical models is expected to have a number of different advantages for the investigative process. The statistical model will not be able to give certainties, but the probabilities, which are calculated, are empirically researched and not based on prejudices. The probabilities of the offender characteristics can be calculated at the beginning of an investigation. The outcome can assist with the decision as to which lines of enquiry need to be explored first. If an offender is likely to live within a 3-km radius of the first encounter site, we might want to start directing our efforts towards the investigation of potential suspects in the neighbourhood, whilst if he is less likely to live within those 3 km, we might want to focus in an earlier stage on a larger area and try to determine if the offender can be tracked down according to the transport he used or focus more on expected criminal record. The resources of the investigative team can be directed to those aspects that are most likely to lead to the offender.

In order to effectively use the statistical models, it is important that all crime characteristics are asked for in the victim statement. If certain elements do not become clear from the statement, it is important to find out if the victim was not able to say anything about this element or if the interviewer mistakenly did not pay attention to this aspect of the crime behaviour.

The following issues might have influenced the results of the predictive models to a certain extent. First, the predictive models in this study were developed from police data that has not collected specifically for the purpose of empirical research. Therefore, the completeness of the data cannot be guaranteed. In addition, the quality of the information registered was not perfectly controlled. Therefore, the results of the study are probably not as reliable and valid as we would have liked. However, this is a general problem in using police data for empirical research. As our aim was to develop a practical instrument for the police organisation, the use of police data was a logical consequence.

Second, the predictive models in this research were developed from data collected from solved rape cases. There is a general weakness regarding the use of this data as a proxy for all offenders, as successful convictions represent only a proportion of all committed offences. As rape has a relatively low conviction rate (Mali, 2007), this is a serious problem. Interpretation of research about offender characteristics based on convicted offenders should be done with caution.

Third, in order to guarantee the independence of cases, only one case per serial offender was included in the sample. The consequence of this solution was that serial offences were underrepresented in the sample compared to the actual situation.

Finally, the predictive value of the models needs to be further tested with 'new offences' that were not used to construct the model. In order to make the models as accurate as possible, all stranger rape cases meeting the criteria mentioned in the creation of the dataset were used for the development of the models. Therefore, in this research, the predictive powers of the models have only been tested based on the same database that was used to construct the model. Further research is needed to determine if the predictive levels of the models will hold when the models are used with new stranger rape cases. One difficulty is that stranger rape is a rare crime; it might be quite some time before sufficient new stranger rape offenders are convicted in the Netherlands. Since we do not know the durability of the model, i.e. to what extent current stranger rapists commit their offences in a similar manner as future stranger rapists will, other possibilities need to be explored as well.

Despite these criticisms, the current statistical models seem to be a promising tool to aid detectives in making decisions about prioritising lines of enquiry in stranger rape offences. Further development and validation of the models may be of significant value for criminal investigation.

## REFERENCES

- Adhami, E., & Browne, D. (1996). *Major crime enquiries: Improving expert support for detectives*. Police Research Group Special Interest Series, Paper 9. London: Home Office.
- Aitken, C. G. G., Connolly, T., Gammerman, A., Zhang, G., Bailey, D., Gordon, R., & Oldfield, R. (1996). Statistical modelling in specific case analysis. *Science and Justice*, 36, 245–255.
- Aitken, C., Connolly, T., Gammerman, A., Zhang, G., & Oldfield, D. (1995). *Predicting an offender's characteristics: An evaluation of statistical modelling*. Police Research Group Special Interest Series, Paper 4. London: Home Office.
- Alison, L., Smith, M. D., Eastman, O., & Rainbow, L. (2003). Toulmin's philosophy of argument and its relevance to offender profiling. *Psychology, Crime and Law*, 9, 173–183.
- Bayley, D. H., & Bittner, E. (1984). Learning the skills of policing. *Law and Contemporary Problems*, 47, 35–59.
- Canter, D. V., Benell, C., Alison, L. J., & Reddy, S. (2003). Differentiating sex offences: behaviorally based thematic classification of stranger rapes. *Behavioral Sciences and the Law*, 21, 157–174.
- Davies, A. (1992). Rapists' behavior: A three aspect model as a basis for analysis and the identification of serial crime. *Forensic Science International*, 55, 173–194.
- Davies, A., Wittebrood, K., & Jackson, J. L. (1998). *Predicting the criminal record of a stranger rapist*. Special Interest Series, Paper 12. London: Home Office Policing and Reducing Crime Unit.
- Eshof, P. van den, & Kleuver, E. E. de. (2000). *Stranger rape*. Paper presented at the European Viclas Conference, Bramshill, August 2000.

- Eshof P. van den, Kleuver E. E. de, Ho Tham, V., & Zwieters, H. (1995). *General profile analysis of sexual assault and rape cases*. Paper presented at the European Viclas Conference, Bramshill, August 2000.
- Eshof, P. van den, & Nierop, N. (2007). Gedraganalyse: daderprofiel in verkrachtingszaken. In A. Ph. van Wijk, R. A. R. Bullens, & P. van den Eshof (Eds.), *Facetten van zedencriminaliteit* (pp. 127–149). 's Gravenhage: Elsevier.
- Francis, B., Barry, J., Bowater, R., Miller, N., Soothill, K., & Ackerley, E. (2004). *Using homicide data to assist murder investigations*. Research Development and Statistics Directorate, Online Report 26/04. London: Home Office.
- Frenken, J., Gijs, L., & Beek, D. van (1996). Seksuele delinquenten in Nederland. Incidentie, kenmerken en behandeling van de plegers. *Maandblad Geestelijke volksgezondheid*, 4, 372–389.
- Frenken, J. Gijs, L., & Beek, D. van (1999). Sexual offender research and treatment in the Netherlands. *Journal of Interpersonal Violence*, 14, 347–371.
- Groth, A. N., Burgess, A. W., & Holstrom, L. L. (1977). Rape: Power, anger and sexuality. *American Journal of Psychiatry*, 134, 1239–1243.
- Häkkinen, H., Lindlöf, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in a sample of Finnish stranger rapes. *Journal of Investigative Psychology and Offender Profiling*, 1, 17–32.
- House, J. C. (1997). Towards a practical application of offender profiling: The RNC's Criminal Suspect Prioritization System. In J. L. Jackson, & D. A. Bekerian (Eds.), *Offender profiling, theory research and practice* (pp. 177–190). Chichester: Wiley.
- Jackson, J. L., Eshof, P. van den, & Kleuver, E. E. de (1997). A research approach to offender profiling. In J. L. Jackson, & D. A. Bekerian (Eds.), *Offender profiling, theory research and practice* (pp. 107–132). Chichester: Wiley.
- Karlsson (1999). Multivariate analysis ('forensiometrics')—A new tool in forensic medicine. Findings on the victim of sharp-force homicide can predict the inter-relationship with the perpetrator. *Forensic Science International*, 101, 33–41.
- Kleuver, E. E. de, Eshof, P. van den, & Zwieters, H. (1996). Waar woont de dader? Verkrachtungen en geografische gegevens. *Modus*, 5, 16–18.
- Knight, R. A. (1999). Validation of a typology for rapist. *Journal of Interpersonal Violence*, 14, 303–330.
- Knight, R. A., & Prentky, R. A. (1990). Classifying sexual offenders the development and corroboration of taxonomic models. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assaults. Issues, theories and treatment of the offender* (pp. 23–52). Plenum Press, New York.
- Kocsis, R. N. (2006). *Criminal profiling. Principles and practice*. Totowa, NJ: Humana Press.
- Kocsis, R. N., Hayes, A. F., & Irwin, H. J. (2002). Investigative experience and accuracy in psychological profiling of a violent crime. *Journal of Interpersonal Violence*, 17, 811–823.
- Kocsis, R. N., Irwin, H. J., Hayes, A. F., & Nunn, R. (2000). Expertise in psychological profiling: A comparative assessment. *Journal of Interpersonal Violence*, 15, 311–331.
- Mali, B. (2007). Geregistreerde zedencriminaliteit in Nederland. In A. Ph. van Wijk, R. A. R. Bullens, & P. van den Eshof (Eds.), *Facetten van Zedencriminaliteit* (pp. 71–97). Den Haag: Reed Business Information.
- Mokros, A., & Alison, J. L. (2002). Is offender profiling possible? Testing the predicted homology of crime scene actions and background characteristics in a sample of rapists. *Legal and Criminal Psychology*, 1, 25–43.
- Righthand, S., & Welch, C. (2001). *Juveniles who have sexually offended. A review of the professional literature*. OJJDP Report. Washington, DC: Department of Justice.
- Rossmo, D. K. (2000). *Geographical profiling*. Boca Raton: CRC Press.
- Safarik, M. E., Jarvis, J. P., & Nussbaum, K. E. (2002). Sexual homicide of elderly females. Linking offender characteristics to victim and crime scene attributes. *Journal of Interpersonal Violence*, 17, 500–525.
- Safarik, M. E., & Jarvis, J. P. (2005). Examining attributes of homicides. Toward quantifying qualitative values of injury severity. *Homicide Studies*, 9, 1–21.
- Santtila, P., Junkkila, J., & Kenneth-Sandnabba, N. (2005). Behavioural linking of stranger rapes. *Journal of Investigative Psychology and Offender Profiling*, 2, 87–103.

- Santtila, P., Laukkanen, M., & Zappalà, A. (2007). Crime behaviours and distance travelled in homicides and rapes. *Journal of Investigative Psychology and Offender Profiling*, 4, 1–15.
- Snook, B., Cullen, R. M., Benell, C., Taylor, J., & Gendreau, P. (2008). The criminal profiling illusion: what's behind the smoke and mirrors? *Criminal Justice and Behavior*, 35, 1257–1276.
- Snook, B., Wright, M., House, J. C., & Alison, L. J. (2006). Searching for a needle in a needle stack: combining criminal careers and journey to crime for suspect prioritization. *Police, Practice and Research*, 6, 419–432.
- Soothill, K., Francis, B., Ackerley, E., & Fligelstone, R. (2002). *Murder and serious sexual assault: What criminal histories can reveal about future serious offending*. Police Research Series, Paper 144. London: Home Office.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Wijk, A. Ph. Van, Mali, S. R. F., Bullens, R. E. R., Prins, L., & Klerks, P. P. H. M. (2006). *Zeden-criminaliteit in Nederland: Delicten en delinquenten nader in beeld gebracht*. Apeldoorn: Politie & Wetenschap.
- Wilson, G., & Alison, L. (2005). Suspect prioritization in the investigation of sex offences: from clinical classification and profiling to pragmatism. In L. Alison (Ed.), *The forensic psychologist's casebook. Psychological profiling and criminal investigation* (pp. 68–89). Cullompton, Devon: Willan Publishing.
- Yarmey, A. D. (1990). *Understanding police and police work*. New York: New York University Press.

Copyright of Journal of Investigative Psychology & Offender Profiling is the property of John Wiley & Sons Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.