Linking commercial burglaries by *modus operandi*: tests using regression and ROC analysis

C Bennell and DV Canter
Department of Psychology, University of Liverpool, Eleanor Rathbone Building, Liverpool, L69 7ZA, United Kingdom
Science & Justice 2002 42 0 – 0 Received 18 September 2001 revised version accepted 23 April 2002

This paper uses statistical models to test directly the police practice of utilising *modus operandi* to link crimes to a common offender. Data from 86 solved commercial burglaries committed by 43 offenders are analysed using logistic regression analysis to identify behavioural features that reliably distinguish between linked and unlinked crime pairs. Receiver operating characteristic analysis is then used to assign each behavioural feature an overall level of predictive accuracy. The results indicate that certain features, in particular the distances between burglary locations, lead to high levels of predictive accuracy. This study therefore reveals some of the important consistencies in commercial burglary behaviour. These have theoretical value in helping to explain criminal activity. They also have practical value by providing the basis for a diagnostic tool that could be used in comparative case analysis.

***Author for correspondence**
Dept of Psychology, Carleton University,
B552 Loeb Building, 1125 Colonel By Drive,
Ottawa, ON, Canada, K1S 5B6

Key words Forensic science, burglary, *modus operandi,* comparative case analysis, logistic regression, ROC analysis.

Introduction

The behavioural linking task, or comparative case analysis (CCA) as it is now commonly called, has the goal of demonstrating that the same offender has committed two or more crimes [1,2]. The task is of particular importance in the absence of a confession, eyewitness testimony or other forensic evidence such as fibres, fingerprints, or DNA. In these cases, behavioural information must be relied upon to link crimes and the task usually involves an examination of what happened at the crime scenes and where the crimes took place. These aspects of the criminal event are popularly regarded as the offender's *modus operandi* (MO) and they have been the subject of limited empirical study.

MO is a rather vague term used in various ways by different police officers and different crime fiction writers. The use of the concept assumes that there will typically be a high degree of similarity between what an offender does in one crime and what he or she does in another. CCAalso assumes that police officers are able to recognise these similarities and use them to make effective investigative decisions. Yet research has shown that linking decisions are often based on the limited, subjective impressions of investigating officers [3], that these impressions often differ from officer to officer [4], and that investigators often perform poorly on tasks like CCA[5].

There is, therefore, value in determining precisely which aspects of offenders'crime scene actions are most often repeated across crimes. This will move the consideration of MO onto a firmer objective footing. Identified areas of behavioural repeatability may also have practical value as a basis for decision support tools in CCA.

Defining the possible decision outcomes in CCA

CCAcan be fruitfully thought of as a diagnostic task similar, for example, to diagnosing cancer in radiology, assessing risk in psychiatry, predicting storms in meteorology, etc. [6]. The central issue is the validity of linking two or more crimes to a common offender. With two possible decisions (linked or unlinked), and two possible realities (actually linked or actually unlinked), there are four potential decision outcomes for the task (Table 1). The goal in studying any system of diagnosis is to increase validity, either by increasing the frequency of correct decisions (hits or correct rejections) or by decreasing the frequency of incorrect decisions (false alarms or misses).

Table 1 Possible decision outcomes in CCA

		Truth: actually linked	Truth: actually unlinked	
Decision:	linked	*a* hits	*b* false alarms	*a+b*
	unlinked	*c* misses	*d* correct rejections	*c+d*
		a+c	*b+d*	*a+b+c+d=N*

Identifying effective linking features

If the criteria available for making a diagnosis form clear-cut categories, then using these categories directly as diagnostic criteria can lead to effective diagnostic decisions [6]. The accurate diagnosis of cancer, for example, can be made in some cases from readily observable features seen in people with cancer that are not present in people without the disease [7,8]. The use of this approach in linking crimes would require the identification of some linking feature, or set of features, reliably associated with crimes committed by the same offender(s) that are not associated with crimes committed by different offenders.

However, research suggests that perfect discriminators are unlikely to be found in the criminal context. Although claims have been made for the existence of such criminal 'signatures' [9,10] there are strong grounds for thinking they are likely to be rare and unlikely to be identifiable for very frequent crimes such as burglary [11]. While some studies have found some degree of behavioural similarity across crimes committed by the same offender [1,2], the extensive literature on offender versatility [12] suggests that high levels of behavioural similarity will not be categorically associated with linked offences. Therefore, it is of value to identify the degree to which features of an offence may help link that offence to others committed by the same offender.

The primary objective of the present study is the identification of predictive accuracy levels for various linking features, by themselves, or in combination with each other.

Identifying appropriate decision thresholds

If categorical diagnostic criteria are not available in CCA, such that their presence or absence indicates the correct decision, then an appropriate decision threshold needs to be established. A decision threshold refers to a cut-off point along a continuum of evidence whereby any value obtained above that point results in a positive decision [6]. In our case, this threshold may correspond to a particular across-crime similarity score that defines how similar two crimes must be before we predict they are linked. According to Swets et al. [6], the general goal is to set decision thresholds in order to "…produce the best balance among the four possible decision outcomes for the situation at hand" (p. 3).

For each of the four decision outcomes in Table 1, conditional probabilities can be estimated from their frequencies, defined as *a*, *b*, *c* and *d*. These estimates refer to probabilities of making certain decisions given, or conditional upon, certain truths. For example, the hit probability, *p*H, indicates the probability of deciding two crimes are linked given that they are in fact linked. This probability is estimated by dividing the number of linked decisions made when the crimes in question are in fact linked, by the total number of crimes that are in fact linked ($a/(a+c)$). The other three conditional probabilities relating to misses (*p*M), false alarms (*p*FA) and correct rejections (*p*CR) are estimated in a similar fashion.

Since the two probabilities of each column in Table 1 add up to

1, only one cell in each column is needed to measure accuracy [6]. The probability of hits and false alarms, *p*H and *p*FA, are the two values that are most commonly used. However, in deciding where to set the decision threshold it is important to recognise that *p*H and *p*FA are related and vary systematically with the exact position of the threshold. As Swets et al. [6] make clear, it is impossible to make the threshold more lenient (e.g., by basing linking decisions on lower across-crime similarity scores) to increase *p*H without also increasing *p*FA. Alternatively, it is impossible to make the threshold more strict (e.g., by basing linking decisions on higher across-crime similarity scores) in order to decrease *p*FA without also decreasing *p*H.

There are a variety of ways to identify appropriate decision thresholds when carrying out CCA. The most effective procedure is to consider the probabilities that linked crimes and unlinked crimes will actually occur (which can also be estimated using the frequencies in Table 1) as well as the costs and benefits associated with incorrect and correct decision outcomes [13]. If

$$p(\text{unlinked}) = \frac{b+d}{N}$$

and

$$p(\text{unlinked}) = \frac{a+c}{N}$$

then multiplying the ratio of probabilities by the ratio of costs (C) and benefits (B) as in

$$\frac{p(\text{unlinked})}{p(\text{linked})} \times \frac{B_{CR} + C_{FA}}{B_H + C_M}$$

will indicate a threshold point that results in optimal decision-making performance. The problem with this approach, however, is that assigning specific costs and benefits to decision outcomes in CCA can be extremely difficult. For example, how does one calculate the cost of arresting an innocent suspect or the benefit of arresting a guilty one?

As an alternative, it is also possible to set an appropriate decision threshold without considering individual costs and benefits, by simply taking their ratio [13]. For example, a police force may decide it is ten times more important to make correct decisions when faced with linked crimes compared to unlinked crimes. This ratio (1/10) can then be substituted into the above formula in place of specific costs and benefits. Perhaps even more realistically, the above formula can be abandoned altogether and an appropriate decision threshold can be set based on some pre-determined limit relating to the rate of false alarms or hits [13]. For example, a police force may decide they do not have the resources to exceed *p*FA=0.20, and therefore this rate will determine what the appropriate decision threshold is.

The secondary objective of the present study is to explore the impact that different decision thresholds have on linking crimes.

Accurately evaluating linking performance

To achieve the objectives of the present study, a procedure is required that can evaluate the diagnostic accuracy of various linking features and the impact of setting different decision thresholds. Evaluating hit and false alarm rates without also examining the effect of setting different thresholds will provide only a partial, and potentially biased, picture of linking validity. Receiver operating characteristic (ROC) analysis assesses these two aspects of linking performance simultaneously, thereby providing a foundation for overall measures of predictive accuracy independent of decision thresholds [6]. Throughout the past two decades, this technique has become the evaluation method of choice for assessing decision-making performance across a wide range of diagnostic settings [6,13,14].

ROC analysis demonstrates how *p*H and *p*FA change for a particular diagnostic feature, or set of features, as decision thresholds are varied from strict to lenient [6]. Both probabilities are calculated across numerous decision thresholds. These probabilities are then plotted on a ROC graph resulting in a concave curve starting from the lower left corner of the graph and ending at the upper right corner, as shown in Figure 1. As Swets and his colleagues [6] explain:

> "At the far lower left [of the graph] both probabilities are near 0, as they would be for a very strict decision threshold, under which the diagnostician rarely makes a positive decision [e.g., that two crimes are linked]. At the far upper right both probabilities are near 1.0, as they would be for a very lenient threshold, under which the diagnostician almost always makes a positive decision. In between the curve rises smoothly, with a smoothly decreasing slope, to represent all of the possible decision thresholds (for a given accuracy)." (p. 6)

The area under a ROC curve, denoted by the symbol *A*, is a measure of diagnostic accuracy for the particular feature(s) that gave rise to that curve [6]. This measure can range in value from 0.5 (indicating chance accuracy) to 1.0 (indicating perfect accuracy). Thus, the area under the ROC curve will be higher as decision-making accuracy increases. Specifically, *A*=1.0 is indicated by a curve that follows the left and upper axes, and *A*=0.5 is indicated by a diagonal line on the ROC graph (referred to as the positive diagonal) going from the lower left corner to the upper right corner.

Our study, therefore, sets out to determine initially if readily available information about burglaries can be shown to provide a statistically significant basis for linking them to a common offender. The next stage is to carry out ROC analyses in order to calibrate the validity of the various criteria used on their own and in combination, and to examine the effects of setting different decision thresholds.

Method

The sample

The present sample of solved serial commercial burglaries was extracted directly from a database of offences housed in one division of a large metropolitan UK police force. The sample consists of two randomly selected crimes from each of 43 serial burglars who committed burglaries between January 1999 and January 2000. For the purpose of the present study, a commercial burglary was defined as any burglary where an offender targeted a commercial property rather than a domestic dwelling. A serial burglar was defined as any offender convicted of two or more commercial burglaries.

There were two primary reasons for selecting just two offences from each burglar's crime series. First, the majority of offenders (55%) included in the entire sample were known to be responsible for just two commercial burglaries. Second, maintaining a constant distribution of offences across offenders ensures that the results will not be biased by undue weighting being given to very prolific offenders who may have displayed particularly high (or low) levels of behavioural similarity across their crimes.

Trained crime analysts coded all of the offence information pertaining to these crimes. However, because the information was entered directly into a database immediately after each crime took place, an assessment of coding reliability was not possible. This potentially weakens the quality of the information utilised, but that weakness is likely to add noise to the data and therefore reduce the chance of any significant patterns emerging. The data has the advantage that it is from genuine police records, collected for statutory and crime management purposes. Any findings from such data, therefore, can claim some important ecological validity and consequent practical relevance.

Potential sources of bias

It should also be noted that the validity of any findings emerging from this study would be limited by biases in the data. One source of potential bias is a result of focusing solely on serial burglaries and not including non-serial burglaries. While some research suggests that non-serial burglaries may actually be quite rare [15], their absence from the present sample will likely bias the results.

Another source of bias arises from the fact that all burglaries examined in this study have been solved. It is possible that solved burglaries are characterised by higher levels of behavioural similarity than unsolved burglaries. Indeed, this may be one of the reasons why solved burglaries are linked in the first place. If this were true, it would limit the extent to which the findings could be generalised to unsolved offences occurring in the same police division.

The data could also be potentially biased because only commercial burglaries were examined, with each offence having been committed within only one police division during a relatively restricted time period. However, preliminary analyses carried out by the authors suggest that the levels of predictive accuracy found in the present study generally exist for residential as well as commercial burglary, across a number of police divisions, during different time periods (though appropriate decision thresholds appear to be more context dependent). Having said this, no claims are being made that the results from the present study can be directly applied to these other contexts. To make such a statement, more detailed studies would obviously be required.

Lastly, relying on police records as the only source of data in the present study could create potential biases. While there is no obvious alternative method for collecting such data, besides the equally biased option of conducting interviews with offenders, it must be acknowledged that police data can be, and often is, inaccurate [16,17].

Selecting linking features

No comprehensive model exists in the published literature that describes the components of burglars' MO. However, drawing on Green et al.'s [13] cluster analysis study and Maguire and Bennett's [18] extensive interviews with offenders, as well as Merry and Harsent's [19] more recent study of burglary, a number of behavioural domains can be identified. These include: (1) entry behaviours (e.g., whether the offender entered through the front door), (2) target selection choices (e.g., whether the offender targeted a filling station), (3) property stolen (e.g., whether the offender stole jewellery) and (4) internal behaviours (e.g., whether the offender consumed food while in the property).

Within the police database used for this study, information pertaining to entry behaviours, target selection choices and property stolen was coded in dichotomous form across all of the offences, indicating the presence or absence of particular crime scene behaviours. This information was extracted from the database for the present study. However, information pertaining to internal behaviours was not coded by the police force, and therefore this aspect of burglary behaviour could not be examined.

In addition to these three behavioural domains, an important fourth aspect of the crimes, involving offender spatial behaviour, was also examined. This information took the form of the distance in kilometres between every pair of burglary locations. The reason for considering this aspect of burglary behaviour is the growing body of literature indicating that many offenders, including burglars, do not travel far to commit their crimes [20–22]. Within the police database used for this study, information pertaining to this aspect of burglary behaviour was available in geo-coded x-y co-ordinates. This information was also extracted from the database for the present study.

Computational procedures

The dependent variable in the present study was whether the same offender or different offenders committed a pair of crimes. The independent variables were all continuous and included: (1) the distance in kilometres between every pair of crimes, and across-crime similarity measures pertaining to (2) entry behaviours, (3) target selection choices and (4) property stolen. Each of these independent variables is based on the premise that a higher degree of behavioural similarity will be exhibited across crimes committed by the same offender. Thus, it was expected that crimes committed by the same offender would be characterised by shorter inter-crime distances and higher across-crime similarity scores for entry behaviours, target selection choices, and property stolen.

Due to the large number of crime pairs that result from a sample of 86 offences, two computer programs were developed to automate the process of calculating measures of behavioural and spatial similarity. The first computer program takes as input a series of dichotomously coded variables pertaining to each of the three behavioural domains. These variables indicate the presence or absence of the specific behavioural features making up these domains. For example, variables related to entry behaviour include such things as ‘entered through front door’ (*yes/no*), ‘entered on ground floor’ (*yes/no*), and ‘used a screwdriver to gain entry’ (*yes/no*). This program then provides as output a similarity measure between every pair of crimes. These similarity measures provide the basis for the subsequent regression and ROC analyses dealing with each behavioural domain.

Jaccard’s coefficient was used as the similarity measure for each of the three behavioural domains. Jaccard’s coefficient is a measure of association that does not take account of joint non-occurrences. In other words, if a particular behaviour is absent across two crimes, the level of similarity between those crimes will not increase. As an example, consider two burglaries that have been dichotomously coded across 17 entry behaviours, where 0 indicates a behaviour that was absent and 1 indicates a behaviour that was present. The pattern of entry behaviours in crime 1 is 00000000000001111 and in crime 2 it is 11000000001111111. If a equals the number of behaviours present in both crimes (1/1), *b* and *c* equal the number of behaviours present in one crime but not the other (1/0 and 0/1), and d equals the number of behaviours absent from both crimes (0/0), Jaccard’s coefficient can be calculated by

$$\frac{a}{a+b+c}$$

Thus, in the above example, where a=4, b=5 and c=0, Jaccard’s coefficient is equal to 0.44 (a value of 1 would indicate total similarity in the behaviours expressed and a value of 0 would indicate no similarity in the behaviours expressed).

Considering the unverifiable nature of burglary data, and the distinct possibility that variables were not recorded as being present when they were in fact present, it may be useful to ignore joint non-occurrences when assessing across-crime similarity. The use of Jaccard’s coefficient for this purpose is also in line with previous examinations of CCA [23], as well as numerous other studies that have utilised police data to identify patterns in offending behaviour [24–27]. However, it should be pointed out that Jaccard’s coefficient is a relatively coarse-grained coefficient and therefore it may be useful in the future to develop a more refined similarity measure.

The second computer program takes as input the geo-coded x-y co-ordinates from the police database and provides as output the distance in kilometres between every pair of crimes. These inter-crime distances provide the basis for the subsequent regression and ROC analyses dealing with distances.

Statistical procedures

Logistic regression analysis was used in the present study to examine the possibility of utilising various linking features to carry out CCA. In this context, the log odds of a crime pair being linked are expressed as a linear combination of across-crime similarity scores. This combination of scores can be expressed in the form of a logistic regression equation, as in

$$\log\left(\frac{p}{1-p}\right) = \alpha+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n$$

where p is the probability of a crime pair being linked, α is a constant, and $\beta_1\ldots\beta_n$ are logit coefficients with which to multiply the observed across-crime similarity scores, represented as $X_1\ldots X_n$.

The log odds, calculated using the above formula, can easily be transformed into the odds of a crime pair being linked, which is a ratio of the probability that a crime pair is linked to the probability that the crime pair is unlinked. To calculate the odds, the log odds are simply exponentiated, as in

$$\text{odds(linked)}=e^{\alpha+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n}$$

If the odds are equal to 1, a crime pair is just as likely to be linked as it is to be unlinked. In contrast, if the odds are less than 1 a crime pair is more likely to be unlinked, and if the odds are greater than 1 a crime pair is more likely to be linked.

The odds can also be converted into a probability that a crime pair is linked. These probabilities are calculated by dividing the odds by 1 plus the odds, as in

$$p(\text{linked}) = \frac{\text{odds}}{1+\text{odds}} = \frac{e^{\alpha+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n}}{1+e^{\alpha+\beta_1X_1+\beta_2X_2+\ldots+\beta_nX_n}}$$

The probability of a crime pair being linked can range from 0 to 1, with higher values indicating a greater chance of being linked.

Two different logistic regression methods were used in the present study. The first method was direct logistic regression where linking features are entered into the regression model simultaneously [28]. This method was used to examine the linking features separately. The second method was forward stepwise logistic regression where linking features are entered into the regression model in a stepwise fashion [28]. As Getty et al. [7] explain, the variable added at each step is the one that, “…most improves the predictive power of the [model] given the set of variables already included” (p. 473). This process stops once the addition of any more variables fails to result in a significant increase in the models predictive power. Forward stepwise logistic regression was used to identify the optimal combination of features for linking purposes.

Statistical issues

When working with log odds, odds and probabilities in logistic regression analysis, it is important to remember two general points. The first point is that all three values provide the same information, only in a slightly different form. Therefore, which values are used is simply a matter of preference. The second point is that all three values are effected by how often linked crime pairs occur. For example, since linked crime pairs will usually be rare compared to unlinked crimes pairs it should come as no surprise when linked crime pairs are associated with relatively low probabilities. What is important in this case are not the actual values of these probabilities, but rather how these probabilities compare to the probabilities associated with unlinked crime pairs.

Also in relation to the use of logistic regression analysis in this study, another important point must be addressed. Typically, the dependent variable used when carrying out regression analysis is statistically independent, in the sense that error associated with one observation is not associated with error from any other observation [29]. This is as it should be. In the present study, however, sampling all possible pairs of crimes consists of observations that may not be statistically independent, since different pairs include crimes committed by the same offender. When the dependent variable is not independent, problems can arise. In such cases, the estimates of standard error corresponding to regression coefficients tend to be smaller than they actually are, though the coefficients themselves will not be biased. This is problematic because it means that inferential tests that depend on these estimates of error cannot be relied upon [30]. Thus, while goodness-of-fit tests will not be problematic in the present study, tests used to measure the predictive accuracy of specific independent variables (e.g., Wald's test) might be.

In this study, the problem of independence is avoided to a large extent because measures of predictive accuracy for each linking feature, or combination of features, are generated from their corresponding ROC curves rather than from regression analysis (see below). The measures of accuracy used in ROC analysis do not rely on estimates of standard error in the same way that formal inferential tests in logistic regression do. As a result, the derived measures of predictive accuracy should not be biased in the way just described even if the dependent variable examined in the present study is not statistically independent.

Evaluating linking performance

In order to reduce the potential bias that exists if regression models are developed and tested on the same sample of commercial burglaries, the present sample was randomly split in half to form an *experimental sample* and a *test sample* [31,32]. Logistic regression models were developed using only the data in the experimental sample and ROC analyses were carried out using only the data in the test sample. The data in the test sample consists of estimated probabilities for each crime pair, calculated using the logistic regression models constructed from the experimental sample. All ROC analyses were carried out using ROCKIT©, a computer package designed by the Department of Radiology at the University of Chicago [33].

This procedure of developing and testing the regression models

Table 2 Summary of logistic regression analyses

Variables	Statistics	Model 1	Model 2	Model 3	Model 4	Model 5 (optimal)
Constant	Logit coeff.	-2.17	-5.08	-5.16	-4.72	-2.82
	Standard error	0.34	0.33	0.43	0.30	0.44
Distance	Logit coeff.	-0.97				-0.88
	Standard error	0.19				0.19
Entry	Logit coeff.		2.68			2.05
	Standard error		0.70			0.73
Target	Logit coeff.			1.98		
	Standard error			0.87		
Property	Logit coeff.				1.33	
	Standard error				0.71	
	Model X^2	43.80	11.80	4.25	3.08	50.88
	Sig. of X^2	p<0.001	p<0.001	p<0.05	p<0.10	p<0.001
	R^2	0.19	0.05	0.02	0.02	0.023

Models 1-4: method of analysis was direct logistic regression
Model 5: method of analysis was forward stepwise logistic regression (inclusion criteria: p<0.05)
Dependent variable: linked crime pair (1), unlinked crime pair (0)

on two separate samples provides some indication of model validity. The degree of validity, however, will depend on how closely the test sample approximates reality and the potential biases previously discussed must again be considered. The solved serial burglaries examined in the present study are probably similar to a portion of commercial burglaries that will occur within this police force in the future. Consequently, it is appropriate for these offences to form part of the test sample. In spite of this, a more realistic test sample would also have included non-serial burglaries as well as unsolved burglaries if this were in fact possible. Since non-serial burglaries are not included in the test sample, the results in the present study should be interpreted with an appropriate level of caution.

Results

Single feature regression models

Direct logistic regression analysis was first run on each of the linking features separately to determine the extent to which single feature regression models can successfully predict whether crime pairs are linked or unlinked. The first four columns in Table 2 contain a summary of these models, including their logit coefficients, standard errors, model X^2 values and R^2 values.

The results in Table 2 suggest that the single feature regression models are able to reliably distinguish between linked and unlinked crimes. This is reflected in the fact that all regression models have X^2 values that are significant at the 10% level (at least) indicating a good degree of fit with the data. However, these X^2 values also suggest that the models differ with respect to their level of fit. Specifically, the model including inter-crime distances appears to be the most accurate, followed respectively by the models including entry behaviours, target selection choices and property stolen. This ordering of the models is also consistent with the R^2 values presented in Table 2, which indicate the proportion of variance explained in the dependent variable by each regression model. Excluding the model containing inter-crime distances, all of the R^2 values are extremely low.

As expected, the signs of the logit coefficients included in the first four columns of Table 2 suggest that crimes pairs committed by the same offender tend to be shorter distances from one another (logit = -0.97) but have higher levels of across-crime similarity for entry behaviours (logit = +2.68), target selection choices (logit = +1.98), and property stolen (logit = +1.33). To determine what these logit coefficients mean in more practical terms, they can be exponentiated. This procedure indicates how a change of ‘c’ units in any of the independent variables affects the odds that two crimes are linked [34].

As an example, the effect of increasing the distance between two crimes by 1.00 km is

$$\text{odds} = e^{(c \times \text{logit})} = e^{(1.0 \times -0.97)} = 0.38$$

which suggests that for every increase of 1.00 km between any two crimes, the odds that the crimes are linked are multiplied by 0.38 (which would reduce them).

Alternatively, the impact of changes in the independent variables can be examined in terms of changes in probability. For example, given the model for distance in Table 2, and a pair of crimes that are 1.00 km apart, the probability that those crimes are linked can be estimated

$$\text{log odds (linked)} = -2.17 - 0.97(1.00) = -3.14$$
$$\text{odds(linked)} = 0.04$$
$$\text{p(linked)} = 0.04$$

Thus, the probability of two crimes being linked when they are 1.00 km apart is 0.04, which is relatively high considering the extremely low percentage of linked crime pairs in the sample.

In contrast, given a pair of crimes that are 2.00 km apart, the estimated probability of the crimes being linked can be seen to decrease by 0.02

$$\text{log odds (linked)} = -2.17 - 0.97(2.00) = -4.11$$
$$\text{odds(linked)} = 0.02$$
$$\text{p(linked)} = 0.02$$

confirming that linked crime pairs do tend to be characterised by shorter inter-crime distances.

Similarly, each of the other logit coefficients in Table 2 can be exponentiated to determine how changes in their values affect the odds that two crimes are linked. In these cases, however, a change of 1 unit will not be particularly meaningful considering that each similarity measure has a potential range from 0 to 1. Instead, it makes more sense to examine the effect of increasing the measures by 0.10, which can be calculated by multiplying the logit coefficients by 0.10 before exponentiating them. Given the three models for entry behaviours, target selection choices and property stolen in Table 2, the effect of increasing the across-crime similarity measures between any two crimes by 0.10 is 1.31, 1.22 and 1.14 respectively. In other words, for every increase of 0.10 units, the odds that the crimes are linked would be multiplied by 1.31 for entry behaviours, 1.22 for target selection choices, and 1.14 for property stolen.

Multiple feature regression models

To determine the extent to which combinations of linking features can successfully predict whether crime pairs are linked or unlinked, forward stepwise logistic regression analysis was used. The last column in Table 2 contains a summary of the optimal regression model. As would be expected, the optimal model contains the two most effective predictors from the previous analysis, which were inter-crime distances and entry behaviours. Also unsurprisingly, the optimal model is able to distinguish between linked and unlinked crimes more accurately than any single feature regression model, as indicated by the significantly higher X^2 value associated with this model as well as the higher R^2 value.

As indicated in Table 3, one reason why target selection choices and property stolen were not included in the optimal model, even though they were relatively accurate as single linking features, is because significant correlations exist between the linking features. The important exception to this is the correlation between inter-crime distances and entry behaviours. As a result, it is highly likely that each linking feature will not *uniquely* account for a significant portion of the variance in the dependent variable, which would enable them all to be included in the optimal model [36]. The correlations presented in Table 4 support this argument. They show that while each linking feature is significantly correlated with the dependent variable, only two features remain highly correlated when the effects of all other features are removed. The remaining two features are inter-crime distances and entry behaviours, which explains why they form the optimal regression model.

Table 3 Correlations between linking features.

	Entry	Target	Property
Distance	-0.03 $p>0.05$	-0.12 $p<0.001$	-0.10 $p<0.001$
Entry		0.16 $p<0.001$	0.18 $p<0.001$
Target		0.14 $p<0.001$	

Table 4 Correlations and partial correlations between linking features and the dependent variable.

	Correlations	Partial correlations
Distance	-0.13 $p>0.05$	-0.12 $p<0.001$
Entry	0.09 $p<0.001$	0.08 $p<0.001$
Target	0.06 $p<0.05$	0.03 p>0.10
Property	0.05 $p<0.10$	0.01 p>0.10

As before, the practical significance of the logit coefficients in this optimal regression model can be explored by exponentiating them. When these calculations are carried out, it can be seen that for every increase of 1.00 km between any two crimes, the odds that the crimes are linked are multiplied by 0.41. When increasing similarity measures pertaining to entry behaviours by 0.10, the odds that the crimes are linked are multiplied by 1.23.

Evaluating linking performance

In order to get some general measure of model validity that indicates how well each model might perform on other solved commercial burglaries in the same police division, the regression models presented in Table 2 were used to calculate estimated probabilities for every possible crime pair in the test sample. These probabilities were then used to construct five ROC graphs, one for each linking feature separately and one for the optimal combination of features. These ROC graphs are presented in Figure 1 along with their overall levels of predictive accuracy as measured by the area under each ROC curve.

Consistent with the analysis of the experimental sample, the ROC curves generated from the single linking features indicate that each feature results in overall levels of accuracy that are significantly greater than chance ($p<0.001$). However, in terms of their predictive accuracy, the ordering of linking features is slightly different than expected from the experimental sample. Clearly, inter-crime distances are still the most accurate linking feature ($A=0.80$), but this is now followed respectively by similarity measures pertaining to target selection choices ($A=0.68$), entry behaviours ($A=0.65$) and property stolen ($A=0.63$). The level of accuracy resulting from the use of inter-crime distances is significantly greater than the levels of accuracy obtained when using target selection choices, entry behaviours or property stolen ($p<0.05$). However, no significant differences were found between these three aspects of burglary behaviour ($p>0.10$).

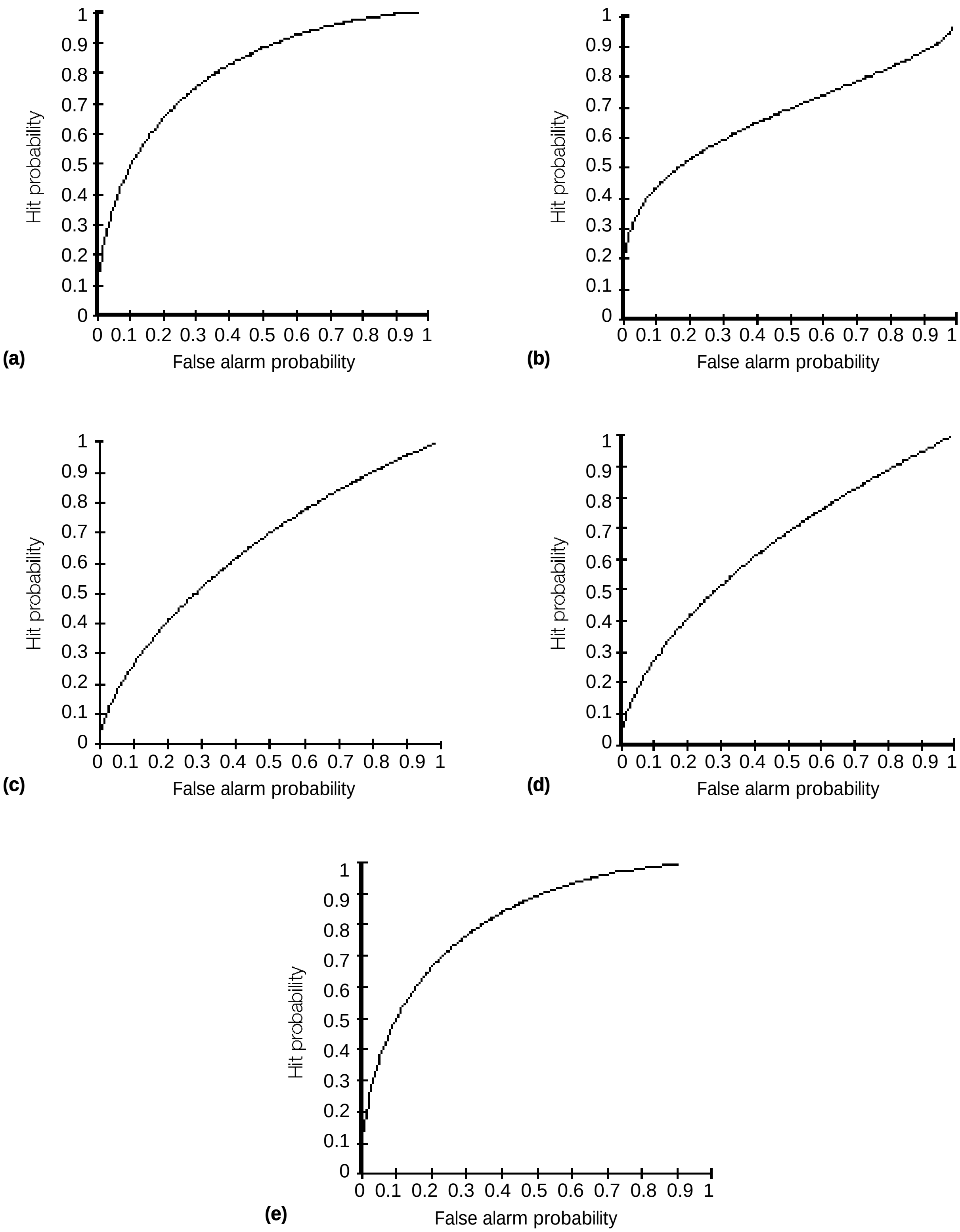
Also consistent with the analysis of the experimental sample, the ROC curve generated from the optimal linking features indicates that this combination of features result in an overall level of accuracy slightly higher than any single linking feature ($A=0.81$). However, the level of overall accuracy obtained when using inter-crime distances and entry behaviours is not significantly greater than the level of accuracy obtained when using inter-crime distances alone ($p>0.10$). This result may seem at odds with the finding in the experimental sample, where the optimal regression model fit the data significantly better than the distance-only model. This can be explained by the fact that the similarity measure pertaining to entry behaviours has a lower level of predictive accuracy in the test sample compared to what they had in the experimental sample.

The impact of setting different decision thresholds

The impact that different decision thresholds have on linking performance is also made clear in the ROC graphs. Regardless of the model used, as decision thresholds are made more lenient, pH and pFA both increase. This can be illustrated in Figure 1 using the ROC curve generated from inter-crime distances. Consider a decision threshold of p≥0.05, which corresponds to an approximate inter-crime distance of 0.70 km. At this particular threshold, 52.4% of linked crime pairs are correctly classified while 93.2% of unlinked crime pairs are correctly classified. However, at the more lenient threshold of p≥0.01, which corresponds to an approximate inter-crime distance of 2.50 km, 61.9% of linked crime pairs are correctly classified while only 67.7% of unlinked crime pairs are.

The practical significance of using different linking features is

Figure 1 ROC graphs for single and optimal linking features. (a) Distance *A*=0.80 (b) Target *A*=0.68 (c) Entry *A*=0.65 (d) Property *A*=0.63 (e) Optimal *A*=0.81.



also made clear in the ROC graphs by considering how many more hits (or how many less false alarms) will be made at a particular decision threshold depending upon the feature selected for analysis. Take the previously mentioned example where a police force decides to set a limit on the rate of false alarms at pFA=0.20. For the ROC curve corresponding to property stolen in Figure 1, this particular threshold results in ROC co-ordinates of pH=0.40 and pFA=0.20. The threshold point for the ROC curve corresponding to inter-crime distances at the same pFA has a pH=0.64. Thus, if an investigator is primarily concerned with making additional hits, they could identify 24 additional linked crime pairs for every 100 pairs encountered if inter-crime distances were drawn on instead of property stolen. Similar comparisons can be made between any of the other linking features.

Discussion and conclusion

Logistic regression and ROC analysis have been used to determine if the degree of across-crime similarity in cases of commercial burglary is high enough for selected aspects of burglary behaviour to allow different crimes to be validly linked to the same offender. Both forms of analysis support the possibility of utilising objectively available aspects of a burglar's MO in a systematic way to carry out valid CCA.

Linking crimes through spatial similarity
It has been demonstrated that the distance between burglary locations is an extremely consistent and stable aspect of commercial burglary behaviour within the particular police division where the present study was carried out. At a theoretical level then, this study adds something to the growing body of literature that has indicated, since the work of White [36] and Shaw [37], that offenders typically do not travel very far to commit their crimes. The results reported here take the understanding of criminal mobility a stage further by indicating that burglars within a given police jurisdiction, who do not live too close to one another, may have relatively distinct areas of criminal activity. This accords well with Grubin et al.'s [2] recent study of rapists where it was shown that linking accuracy could be enhanced by using spatial information. However, Grubin et al.'s study suffered from the possible artefact that the rape series they examined were drawn from all over the UK. Distinguishing such offence series by the locality in which the offences occurred is therefore not as stringent a test as the present study of commercial burglaries where a relatively small area of the country was examined.

In practical terms, this finding suggests that inter-crime distances can provide a powerful, and relatively simple way of linking commercial burglaries. Specifically, this linking feature may prove useful as an effective first filter when carrying out CCAto reduce the number of potential links that initially need to be examined. Considering the current state of technology in the majority of modern police forces, it would be feasible to draw on the results presented in this paper and combine them with digital maps of police divisions to create likely 'linkage areas'. Crimes committed within certain distances of one another could be given a high priority and then additional analytical techniques could be used to further reduce the number of false alarms. Drawing on different burglary behaviours, obvious signature aspects, temporal information, or police intelligence could prove useful for this purpose.

Linking crimes through behavioural similarity
The present study also indicates that similarity measures pertaining to other behavioural domains can be used to link commercial burglaries, though not to the same degree as inter-crime distances. Consequently, the validity of CCA, in its initial stages at least, will depend on what features are used to perform the analysis. The lower level of predictive accuracy for target selection choices, entry behaviours and property stolen is generally consistent with existing research. This research suggests that crime scene behaviours often change across crimes due to external situational influences and internal learning processes [38,39]. Nevertheless, the police often use these behaviours for linking purposes, either formally or informally, and in some cases they form the basis for a legal argument of similar fact evidence [40]. Therefore, there is some value in assessing the accuracy of each behavioural domain in order to understand the patterns of activity that burglars exhibit.

While similarity measures pertaining to target selection choices, entry behaviours and property stolen provide a basis for linking crimes committed by the same offender, their level of predictive accuracy (relative to one another) appears to vary across different samples of offences. This finding supports the idea that these behaviours are more context-dependent than inter-crime distances. Indeed, the results presented in this paper provide preliminary evidence that the differences in predictive accuracy levels across all linking features may relate to how situation-dependent the features are. The property an offender steals, for example, is perhaps the most situation-specific set of behaviours in burglary, depending as they do on what is available to be stolen. The recording of this information in official records may also be unreliable both because of what the police choose to record and because of what the property owner chooses to say was stolen [16]. This may explain why property stolen leads to the least accurate predictions in both the experimental sample and the test sample.

In general, as linking validity increases so to does the apparent extent to which linking features consist of behaviours that can be determined by the offender – from the property they steal, to their entry and targeting behaviour, to where they initially go to commit the crime. This finding is consistent with studies of non-criminal consistency, where operant behaviours (i.e., behaviours emitted by the person across a range of situations) are usually exhibited in a more consistent fashion than respondent behaviours (i.e., behaviours that require specific, eliciting stimuli within situations) [41,42]. The practical importance of such a finding is that it may provide investigators with a means of predicting, *a priori*, what aspects of burglary behaviour will be most useful for CCA.

Combining linking features to enhance linking performance
Another significant finding in the present study is the possibility

of achieving higher levels of accuracy in CCA when combinations of carefully chosen features are used. Compared to the single feature models, an increase in the overall level of predictive accuracy was observed in the experimental sample and the test sample when inter-crime distances and entry behaviours were used simultaneously. However, in the test sample this increase did not reach the point of being statistically significant compared to the distance-only model. One of the reasons for this was that entry behaviours had less predictive power in the test sample. This reinforces the need to identify stable linking features when developing optimal linking models; features that maintain a high level of predictive accuracy across different samples of offences. If such features can be identified in commercial burglary, beyond those related to spatial behaviour, then combining these features with the distances that exist between burglary locations will likely result in models with significantly more predictive power.

Despite the lack of a significant finding in the test sample, examining the accuracy of feature combinations is important for CCAbecause of the inherent unreliability in any single piece of information collected as part of a police investigation. A careful combination of selected features could counteract problems there might be in recording such material. Discovering a way of achieving maximum predictive power in CCA using the fewest possible number of linking features is also important because it would reduce the need to collect a great deal of information on a crime, with the attendant problems of such large-scale data collection. It implies that collecting appropriate, possibly limited, information carefully may be more effective than collecting a great deal of information in the hope that some of it may turn out to be of value. Thus, instead of developing longer, more comprehensive linking pro forma's, the method of analysis presented in this study open up the possibility of finding ways to provide more manageable guidance that is just as effective, simply by cutting out unneeded redundancies in the behavioural features that are used.

The importance of decision thresholds
Lastly, the present study demonstrates the impact that decision thresholds have on linking performance. When using single linking features or combinations of features, linking accuracy was shown to depend on the exact position of the decision threshold. Specifically, both *p*H and *p*FA could be seen to increase in value as decision thresholds became more lenient. This highlights the need to identify appropriate decision thresholds in CCA that produce a desired balance between the four possible decision outcomes.

One possible strategy for accomplishing this goal was examined here, whereby a limit was set on the rate of false alarms that could be made. However, alternative strategies also exist. These alternatives require the costs and benefits associated with the various decision outcomes in CCA to be made explicit. Such decisions may be extremely difficult to make, in particular when human rights and lives are at stake, and the decisions will necessarily involve both economic as well as ethical considerations. However, carrying out such cost-benefit analyses could lead to decision support tools in CCA that are fine-tuned to quite specific investigative situations, in a similar way to what is being done in other diagnostic settings [43].

Acknowledgements
This research was supported by an Overseas Research Scholarship awarded to the first author by the Overseas Research Students Awards Scheme. The authors would like to thank Chris Setzkorn for his assistance with software development. The authors are also grateful to Brent Snook, Laurence Alison and two anonymous reviewers for their helpful comments on an earlier version of this article.

References

1 **Green EJ, Booth CE and Biderman MD.** Cluster analysis of burglary MO's. Journal of Police Science and Administration 1976; 4: 382-387.

2 **Grubin D, Kelly P and Brunsdon C.** Linking Serious Sexual Assaults Through Behaviour. London: Home Office, 2001.

3 **Canter DV.** Criminal Shadows: The Inner Narratives of Evil. Texas: Authorlink Press, 2000.

4 **Maltz MD, Gordon AC and Friedman W.** Understanding and using information about crime. In: Maltz MD, Gordon AC and Friedman F, eds. Mapping Crime in its Community: Event Geography Analysis. New York: Springer-Verlag, 1990: 44–58.

5 **Wilson M, Canter DV, Jack K and Butterworth D.** The Psychology of Rape Investigations. ESRC Report, University of Liverpool, Liverpool, 1996.

6 **Swets JA, Dawes RM and Monahan J.** Psychological science can improve diagnostic decisions. Psychological Science in the Public Interest 2000; 1: 1–26.

7 **Getty DJ, Seltzer SE, Tempany CMC, Pickett RM, Swets JA and McNeil BJ.** Prostate cancer: Relative effects of demographic, clinical, histologic, and MR imaging variables on the accuracy of staging. Radiology 1997; 204: 471–479.

8 **Lo JY, Baker JA, Kornguth PJ, Iglehart JD and Floyd CE.** Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features. Radiology 1997; 203: 159–163.

9 **Keppel RD and Birnes WJ.** Signature Killers: Interpreting the Calling Cards of the Serial Murderer. New York: Pocket Books, 1997.

10 **Keppel RD.** Signature murders: A report of the 1984 Cranbrook, British Columbia cases. Journal of Forensic Sciences 2000; 45: 508–511.

11 **Canter DV.** Offender profiling and criminal differentiation. Legal and Criminological Psychology 2000; 5: 23–46.

12 **Klein MW.** Offence specialisation and versatility among juveniles. British Journal of Criminology 1984; 24: 185–194.

13 **Swets JA.** Signal Detection Theory and ROC Analysis in Psychology and Diagnostics. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

14 **Swets JA and Pickett RM.** Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York: Academic Press, 1982.

15 **Hindelang MJ, Hirschi T and Weis JG.** Measuring Delinquency. Beverley Hills, CA: Sage Publications, 1981.

16 **Alison LJ, Snook B and Stein K.** Unobtrusive measurement: Using police information for forensic research. Qualitative Research 2001; 1: 241–254.

17 **Farrington DP and Lambert S.** The Feasibility of a Statistical Approach to Offender Profiling: Burglary and Violence in Nottinghamshire. London: Home Office, 1992.

18 **Maguire M and Bennett T.** Burglary in a Dwelling: The Offence, the Offender and the Victim. London: Home Office, 1982.

19 **Merry S and Harsent L.** Intruders, pilferers, raiders and invaders: The interpersonal dimension of burglar y. In: Canter DV and Alison LJ, eds. Profiling Property Crimes. Aldershot: Ashgate Publishing Limited, 2000: 33–56.

20 **Canter DV and Larkin P.** The environmental range of serial rapists. Journal of Environmental Psychology 1993; 13: 63–69

21 **Rossmo DK.** Target patterns of serial murderers: A methodological model. American Journal of Criminal Justice 1993; 17: 1–21.

22 **Wiles P and Costello A.** The Road to Nowhere: The Evidence for Travelling Criminals. London: Home Office, 2000.

23 **Canter DV and Heritage R.** A Facet Approach to Offender Profiling. London: Home Office, 1991.

24 **Bennell C, Alison LJ, Stein K, Alison E and Canter DV.** Sexual offences against children as the abusive exploitation of conventional adult-child relationships. Journal of Social and Personal Relationships 2001; 18: 149–165.

25 **Canter DV and Fritzon K.** Differentiating arsonists: A model of firesetting actions and characteristics. Legal and Criminological Psychology 1998; 3: 73–96.

26 **Canter DV, Hughes D and Kirby S.** Paedophilia: Pathology, criminality or both? The development of a multivariate model of offence behaviour in child sexual abuse. Journal of Forensic Psychiatry 1998; 9: 532–555.

27 **Salfati CG and Canter DV.** Differentiating stranger murders: Profiling offender characteristics from behavioural styles. Behavioural Sciences and the Law 1999; 17: 391–406.

28 **Tabachnick BG and Fidell LS.** Using Multivariate Statistics. New York: Harper Collins Publishers, 1996.

29 **Lewis-Beck MS.** Applied Regression: An Introduction. London: Sage Publications, 1980.

30 **Chatterjee S and Price B.** Regression Analysis by Example. New York: John Wiley and Sons, 1977.

31 **Efron B.** The Jackknife, the Bootstrap and Other Re-sampling Plans. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1982.

32 **Gong G.** Cross-validation, the jackknife, and the bootstrap: Excess error estimation in forward logistic regression. Journal of the American Statistical Association 1986; 81: 108–113.

33 **Metz CE, Herman BA, and Roe CA.** Statistical comparison of two ROC curve estimates obtained from partially paired data sets. Medical Decision Making 1998; 18: 110–121.

34 **Hosmer DW and Lemeshow S.** Applied Logistic Regression. New York: John Wiley and Sons, 1989.

35 **Cohen J and Cohen P.** Applied Multiple Regression / Correlation Analysis for the Behavioural Sciences. London: Lawrence Erlbaum Associates, 1983.

36 **White RC.** The relation of felonies to environmental factors in Indianapolis. Social Forces 1932; 10: 459–467.

37 **Shaw C.** Juvenile Delinquency and Urban Areas. Chicago: University of Chicago Press, 1942.

38 **Douglas JE and Munn C.** Violent crime scene analysis: *Modus operandi*, signature and staging. FBI Law Enforcement Bulletin 1992; 61: 1–10.

39 **Davies A.** Rapist’s behaviour: A three-aspect model as a basis for analysis and the identification of serial crime. Forensic Science International 1992; 55: 173–194.

40 **Ormerod D.** Criminal profiling: Trial by judge and jury, not criminal psychologist. In: Canter DV and Alison LJ, eds. Profiling in Policy and Practice. Aldershot: Ashgate Publishing Limited, 1999: 207–261.

41 **Funder DC and Colvin CR.** Explorations of behavioural consistency: Properties of persons, situations, and behaviours. Journal of Personality and Social Psychology 1991; 60: 773–794.

42 **McClelland D.** Is personality consistent? In: McClelland D, ed. Motives, Personality and Society. New York: Praeger, 1984: 185–211.

43 **Schwartz JS, Dans PE and Kinosian BP.** Human immunodeficiency virus test evaluation, performance, and use: Proposals to make good tests better. Journal of the American Medical Association 1988; 259: 2574–2579.