BEHAVIOURAL CONSISTENCY AND DISCRIMINATION IN SERIAL BURGLARY

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by Craig Bennell

Department of Psychology

June 2002

ACKNOWLEDGEMENTS

I would first like to say thank you to my supervisor, David Canter. I have greatly appreciated his many helpful comments as well as the numerous opportunities he has given me since my arrival in Liverpool.

I would also like to express my appreciation to Laurence Alison who was always willing to lend an ear when needed.

Many of my fellow students have also spent time listening to me try and sort through various problems. In particular, I would like to thank Brent Snook for the many hours he spent discussing research with me, and the time he spent proofreading chapters. I would also like to say thank you to Chris Setzkorn for his valuable assistance with software development.

I must also extend my thanks to various individuals in police forces around the UK for providing me with access to data and for being available when I needed assistance. I would especially like to thank Ian Oldfield in London, Simon Merry and John Shewbridge in Dorset, Sam Frenz and Karen Kelly in Manchester, and Anita Pattenson in Merseyside.

A number of other people provided long distance email assistance to me in times of need. The assistance provided by John Swets, Robyn Dawes, and Charles Metz was very much appreciated.

I would also like to acknowledge the Overseas Research Awards Scheme and the University of Liverpool who provided the funding for my research.

Last, but not least, I would like to thank Cindy, for her love, encouragement and support over the last 5 years.

ABSTRACT

This thesis examines whether, and under what conditions, serial burglars display consistent patterns of behaviour across the crimes they commit. This thesis also examines the extent to which these consistencies, if they do in fact exist, can be used to distinguish between crimes committed by different serial burglars. The extent to which burglars are behaviourally consistent is viewed as a product of how similar the behaviours are that burglars exhibit across their own crimes and how distinct these behaviours are when compared to other offenders committing similar sorts of crimes. If burglars do not exhibit relatively high levels of similarity and distinctiveness, their behaviour cannot be defined as behaviourally consistent. The extent to which burglaries committed by different offenders can be distinguished from one another is viewed as a product of how consistent serial burglars are. In other words, the higher the level of behavioural consistency that exists in a given sample of burglars, the higher the level of discrimination that will be possible within that sample.

To examine these issues, a procedure is proposed based on receiver operating characteristic (ROC) analysis. This procedure is shown to relate fundamentally to the behavioural processes outlined above. As a result, ROC analysis provides a way of accurately quantifying the extent to which serial burglars are consistent across the crimes they commit and the level of discrimination that is possible as a result. The input to the ROC procedure essentially consists of measures of behavioural similarity, which are calculated across pairs of crimes committed by the same or different offenders. Across-crime similarity scores based on any aspect of burglary behaviour can be used to construct an empirical ROC curve, where the height of the curve corresponds to how consistently the behaviour is expressed. Each point falling along a ROC curve relates to a different across-crime similarity score, used as a decision threshold for deciding when two crimes are linked. These points can therefore be used to examine the level of discrimination accuracy possible at each and every decision threshold. The proposed ROC procedure is tested across six samples of solved serial burglary data collected from various police forces across the UK. With some exceptions, the data consists of information about where each burglary took place, as well as information relating to entry behaviours, target selection choices, internal behaviours and property stolen. The ROC results clearly indicate that levels of behavioural consistency depend on what aspect of burglary behaviour is examined. While the majority of behaviours result in very low levels of consistency, one aspect of spatial behaviour is exhibited in an extremely consistent fashion. Specifically, inter-crime distances are found to result in very high ROC curves indicating that the offending areas utilised by different burglars do not overlap to any great extent. This is found to be the case for both residential and commercial burglars who commit crimes in rural or urban areas, regardless of whether the precision of analysis is at the level of an entire police force or smaller police districts.

Consistent with these findings, inter-crime distances are found to result in very high levels of discrimination accuracy compared to all other aspects of burglary behaviour. However, regardless of what behaviour is examined, discrimination accuracy relates directly to the position of the decision threshold. With respect to inter-crime distances, optimal decision thresholds are found to be sample specific but somewhat predictable. For example, to get the same level of discrimination accuracy in commercial burglary, more lenient thresholds have to be used. This is primarily because commercial burglaries tend to be characterised by slightly larger inter-crime distances than residential burglaries. To show that the ROC procedure has predictive power, a number of blind discrimination trials are carried out where optimal thresholds are used to distinguish between crimes committed by different offenders. In every case, the levels of discrimination accuracy achieved across these trials generally correspond to the levels of accuracy predicted by the ROC results.

Possible psychological, methodological and practical explanations for these findings are offered, and their theoretical and practical implications are discussed. It is suggested that the ROC framework has theoretical value in helping to reveal some of the important consistencies in burglary behaviour. It is also suggested that the ROC framework has practical value by providing the basis for a diagnostic tool that could be used to distinguish between crimes committed by different offenders. Areas of possible future research are also presented, ranging from attempts to use the proposed ROC framework to explore general psychological issues in the non-criminal context, to further studies designed specifically to clarify some of the ROC results presented in the current research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
CONSISTENCY AND DISCRIMINATION IN SERIAL BURGLARY

1.1. Introduction

This research examines whether, and under what conditions, serial burglars display consistent patterns of behaviour across the crimes they commit. This research also examines the extent to which these consistencies, if they do in fact exist, can be used to distinguish between crimes committed by different burglars. In order to examine these issues, two things are necessary. First, operational definitions of consistency and discrimination are required. For example, it must be clear from a definition of consistency how burglary behaviour needs to be expressed in order to conclude that a burglar is in fact behaving in a consistent fashion. Second, an effective analytical procedure for examining consistency and discrimination is required. This procedure must not only provide a way of accurately measuring the extent to which consistency exists, it must also measure the degree of discrimination that is possible as a result.

Few attempts have been made to empirically examine behavioural consistency and discrimination in the criminal context. For example, until very recently only one published study had empirically examined the hypothesis that crimes committed by different offenders can be distinguished from one another based solely on an analysis of offence behaviours (Green, Booth & Biderman, 1976). Likewise, only one published study has ever empirically examined the hypothesis that police investigators can use offence behaviours to accurately discriminate between crimes committed by different offenders (Canter & Heritage, 1991). This almost total lack of research is surprising for a variety of reasons, not least of which is the practical importance in the investigative context of establishing whether an offender has committed more than one crime (Wilson, Canter, Jack & Butterworth, 1997).

1.2. What existing studies say about consistency and discrimination

Studying behavioural consistency and discrimination in the criminal context usually involves an examination of what happened at a crime scene and where the crimes took place. These aspects of the criminal event are popularly regarded as the offender's modus operandi (MO). The use of this term generally assumes that there will be a high degree of behavioural similarity between what an offender does in one crime and what he or she does in another, and that these behaviours will be more characteristic of the offender committing the crimes than any other offender committing similar sorts of crimes.

Many researchers have questioned the notion that offenders will exhibit stable patterns of behaviour across the crimes they commit (Davies, 1992; Douglas & Munn, 1992; Turvey, 2000). These researchers point out something that has been known in non-criminal contexts for some time – that an individual's behaviour can change across situations due to a wide variety of internal and external factors. However, despite the range of possible factors that may influence an offender's behaviour, including natural maturation, learning processes, drug use and various situational factors, evidence exists that stable MOs can be identified and be of practical use. Indeed, the two published studies that have empirically examined the degree to which consistency and discrimination exists in the criminal context support the idea that offenders exhibit characteristic behavioural styles to some extent across the crimes they commit (Green *et al.*, 1976; Grubin, Kelly & Brunsdon, 2001).

1.2.1. The Green, Booth and Biderman (1976) study

The first of these two studies was Green *et al.*'s (1976) innovative examination of residential burglary behaviour through the use of cluster analysis. In this study, Green and his colleagues were able to draw on various aspects of burglary behaviour in order to discriminate between crimes committed by different offenders. Specifically, 14 out of 15 solved burglaries, which were committed by 3 different serial burglars, could be accurately linked to the correct offender by considering the location of the targeted dwellings, entry behaviours, the type of property stolen, and the value of property stolen.

While this study provides some evidence that offenders exhibit stable patterns of behaviour across the crimes they commit, and that these behaviours can be used to accurately discriminate between crimes committed by different offenders, some caution is warranted when interpreting the results. The first problem in this study is the extremely low sample size. It seems very unlikely that the high level of discrimination accuracy achieved by Green and his colleagues would be found if more than 15 burglaries had been examined in their study. The second problem is the crime scene behaviours that were used in the analysis. The authors themselves say that they specifically chose 3 known burglars who had clearly defined MOs, making it difficult to see how one would go about identifying discriminatory behaviours when dealing with unsolved crimes in a practical setting. The third problem is how the researchers decided when crimes were linked. There is nothing in Green *et al.*'s (1976) study to indicate how similar two burglaries must be before they should be linked.

1.2.2. The Grubin, Kelly and Brunsdon (2001) study

In a much more recent study, Grubin *et al.* (2001) also provide evidence in support of MO, though this time in relation to sexual assaults. The analytical approach adopted in this study is different from the approach taken by Green *et al.* (1976), but the hypothesis being tested was exactly the same. Essentially, the approach was to reduce sexual assault behaviour to 4 underlying behavioural domains (labelled control, sex, escape and style) and to derive 4 domain types from each of these domains that contained different crime scene behaviours. Grubin and his colleagues were then interested in testing the extent to which offenders display the same domain types across the crimes they commit and the extent to which these domain types allow crimes committed by different offenders to be distinguished from one another.

These researchers were able to draw on various aspects of sexual assault behaviour to demonstrate that sex offenders do behave in a somewhat similar fashion across many of their crimes. Based on a sample of 468 sex offences committed by 210 offenders in the UK, it could be shown that approximately 83% of the offenders displayed at least one domain type in a similar fashion across their entire crime series. For example, a number of offenders repeatedly exhibited behaviours from a domain type labelled control type 4, which consists of the offender planning the offence, attacking indoors, bringing a weapon to the scene and using a surprise attack. Furthermore, it could be shown that about 26%

of the offenders displayed all four of their assigned domain types in a similar fashion across at least two of their crimes.

To determine whether it was possible to use these domain types to discriminate between crimes committed by different offenders, each crime was extracted from the sample one at a time. The 10% of offences most similar to the extracted crime were then examined to determine how many of them were part of the offence series. For series of various lengths, the number of matched offences was significantly greater than chance. Grubin and his colleagues also showed that discrimination performance could be improved by considering when and where the offences took place. For example, by eliminating offences within the 10% sample that were far apart from one another either in space (e.g., >30 km) or in time (e.g., >250 days), discrimination accuracy could be significantly increased.

While this study usefully builds on the research carried out by Green *et al.* (1976), there are a number of problems with it. The first problem is that it is unclear why the discriminators are used in the order they are. In other words, why are behavioural domains initially used to discriminate between offences and then spatial and temporal behaviours instead of the other way around? The second problem is that it is impossible to know from the results of this study how much each domain contributes to the level of discrimination accuracy, if they in fact do. For example, behaviours in different domains may be highly correlated with each other and therefore using all the domains to discriminate between crimes may be somewhat redundant. The third problem is the same problem as mentioned above, which is about how the researchers decided when two crimes were similar enough to warrant being linked. Grubin and his colleagues define the level of similarity required in their study (i.e., the 10% of offences most similar to the index offence), but a different level of similarity could have resulted in higher levels of accuracy.

1.2.3. The Canter and Heritage (1991) study

Given the evidence provided by Green *et al.* (1976) and Grubin *et al.* (2001) for the existence of relatively stable MOs, it might be possible for police investigators to discriminate between crimes committed by different offenders.

Indeed, it is often assumed that police investigators will be able to recognise the important aspects of an offender's behaviour and then use these behaviours to make effective investigative decisions (Craik & Patrick, 1994; Swanson, Chamelin & Territo, 1996). In contrast to this assumption, however, research has shown that such decisions are often based on the limited and subjective impressions of investigating officers (Wilson *et al.*, 1997), that these impressions often differ from investigator to investigator (Maltz, Gordon & Friedman, 1990), and that investigators often perform poorly on such tasks (Canter & Heritage, 1991).

For example, in a simplified version of a real investigative task, Canter and his colleagues provided police investigators with a description of the crime scene behaviours exhibited by 3 different serial rapists across 4 of their crimes (Canter & Heritage, 1991). The task for the investigators was to identify the set of behaviours that distinguished between the crimes committed by different offenders and to make a decision about which of the 12 crimes belonged to each offender. Not only did the investigators disagree about what behaviours were important for the task, they also disagreed about what crimes were actually linked. The majority of investigators performed at a level that was no greater than chance. Indeed, even when investigators did select behaviours that could accurately discriminate between the crimes they often made incorrect decisions, suggesting that they were unable to effectively process the information they selected.

Studies such as this suggest there may be real value in developing analytical procedures to assist with this task. From the studies carried out by Green *et al.* (1976) and Grubin *et al.* (2001), it seems clear that such a procedure would need to perform at least two separate functions. First, the procedure will need to identify the sort of evidence that should be used to make discrimination decisions in the first place. For example, the procedure should allow a user to answer questions such as: What crime scene behaviours are best for predicting whether the same burglar has committed a number of crimes? Second, the procedure will need to identify how much of this evidence is required before a decision should be made that crimes are linked. For example, the procedure should allow a user to answer questions such as: If entry behaviours are better than internal behaviours at discriminating between crimes, how similar do entry behaviours have to be across crimes in order to decide that they are linked?

1.3. Defining consistency and discrimination in the criminal context

In order to develop such a procedure, it is first necessary to define what is meant by consistency and discrimination in serial burglary. In the present research, the extent to which burglars are consistent is viewed as a product of how similar the behaviours are that burglars exhibit across their own crimes and how distinct these behaviours are when compared to the behaviours exhibited by other burglars. In the present research, if burglars do not exhibit similarity and distinctiveness, their offending behaviour cannot be defined as behaviourally consistent. The extent to which burglaries committed by different offenders can be distinguished from one another is viewed as a product of how consistent serial burglars are. The higher the level of consistency exhibited by a sample of burglars, the higher the level of discrimination that is possible as a result.

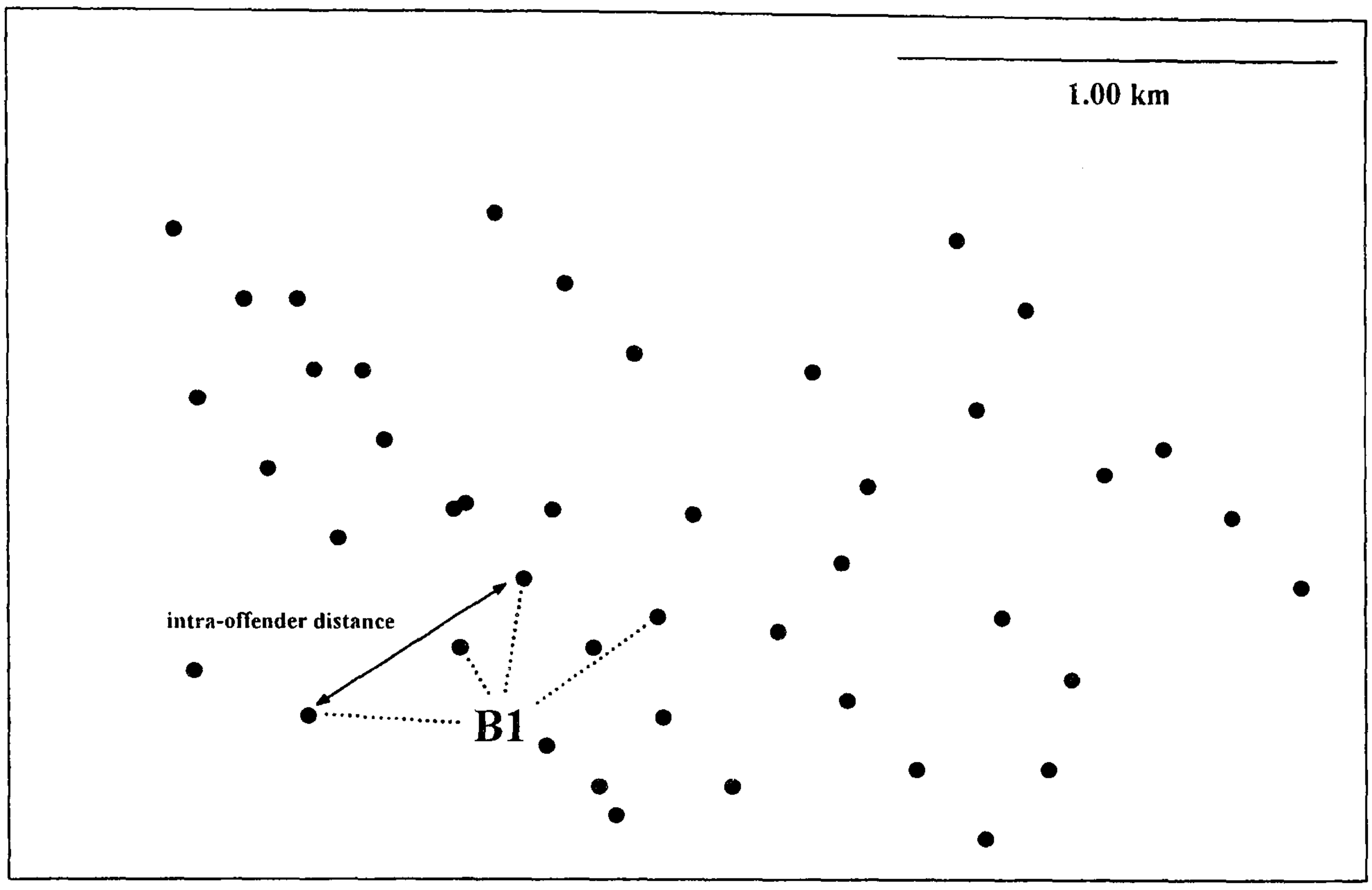
One of the first challenges encountered when developing operational definitions of these various behavioural processes is that while each of the processes are obviously different from one another they are also not totally independent. As a result, behavioural similarity, distinctiveness, consistency and discrimination need to be defined clearly and separately. However, they also need to be presented within an overall framework that indicates how they relate to and depend on one another. Once such a framework is developed, it should be easier to construct an effective analytical procedure for reliably, accurately and objectively measuring the extent to which each process exists. Only when this is done, can the theoretical and practical implications of any findings be thoroughly explored.

1.3.1. Behavioural similarity in serial burglary

Behavioural similarity can be defined as the extent to which burglars behave in a similar fashion across the burglaries they commit, regardless of what burglary behaviour is examined or how it is measured. For example, similarity would be said to exist in the present research if a burglar exhibits similar entry behaviours across each of his burglaries, but it would also exist if a burglar commits his burglaries in a similar geographic area. In the first instance, similarity could be measured by counting the number of similar entry behaviours exhibited across crimes committed by the same offender, where higher scores would indicate higher levels of similarity. In the second instance, similarity could be measured by calculating the distance between crimes committed by the same offender, where lower scores would indicate higher levels of similarity.

As an example of what behavioural similarity might look like in the case of crime site selection, consider the map in Figure 1.1. This map represents all possible burglary targets in a particular geographic area as well as the spatial behaviour of one serial burglar, B1, who has committed four crimes. Consistent with the journey to crime research, which suggests that many burglars travel short distances from home to carry out their crimes (e.g., Baldwin & Bottoms, 1976; Rengert & Wasilchick, 2000; Wiles & Costello, 2000), this burglar never travels further than 0.50 km from home. The fact that this serial burglar exhibits a high level of similarity with respect to his crime site selection is indicated by the existence of relatively short intra-offender distances (an example of which is indicated by the solid arrow). If the burglar in this case exhibited longer intra-offender distances, he would be exhibiting lower levels of similarity.

Figure 1.1. Behavioural similarity (crime site selection)

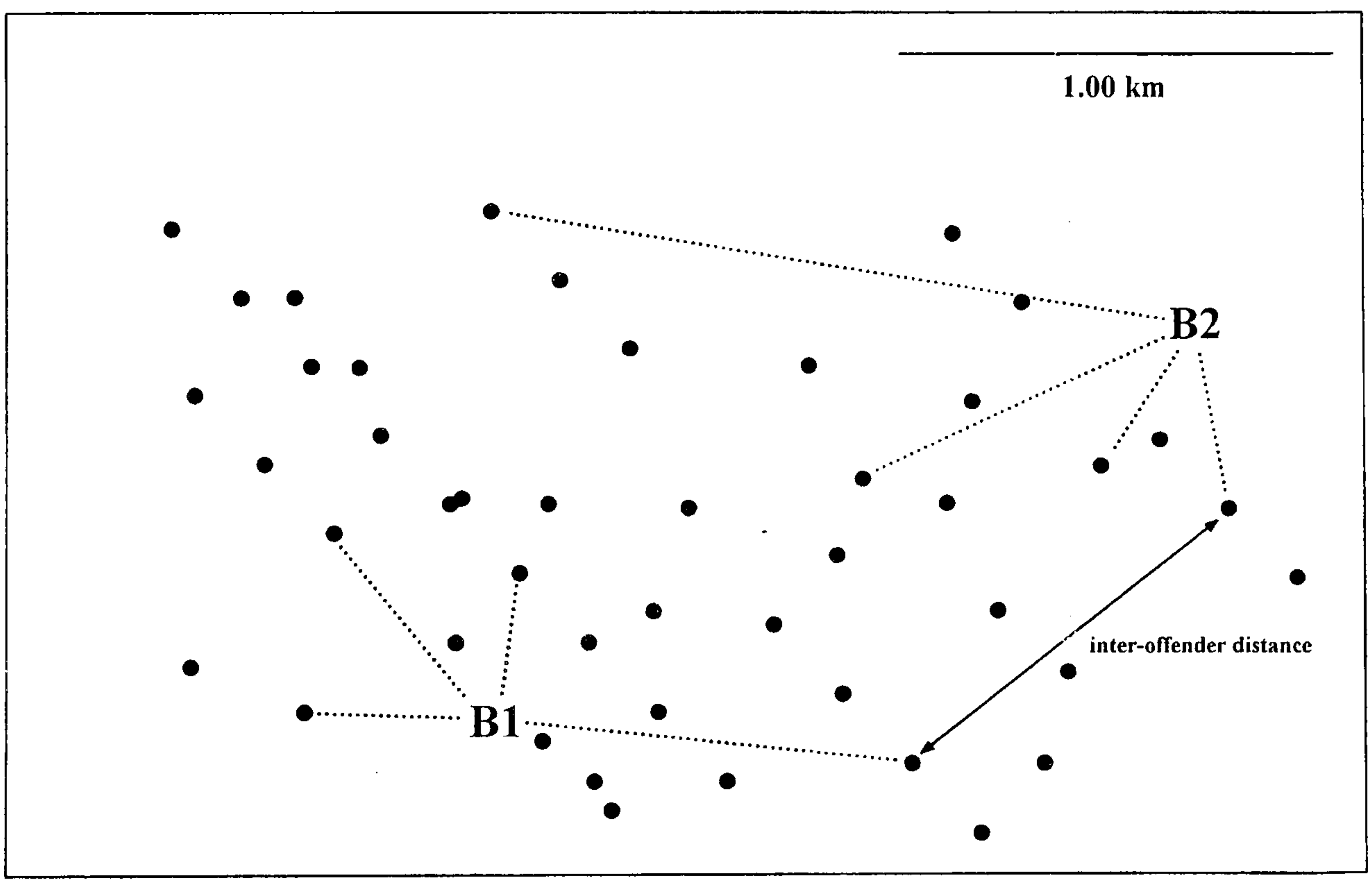


1.3.2. Behavioural distinctiveness in serial burglary

Behavioural distinctiveness can be defined as the extent to which burglars exhibit behaviours across their burglaries that are more typical of them than of other burglars committing similar sorts of crimes, regardless of what burglary behaviour is examined or how it is measured. For example, distinctiveness would be said to exist in the present research if two or more burglars exhibit different entry behaviours across their burglaries compared to one another, but it would also exist if two or more burglars commit their burglaries in different geographic areas. In the first instance, distinctiveness could be measured by counting the number of similar entry behaviours exhibited across crimes committed by different burglars, where lower scores would indicate higher levels of distinctiveness. In the second instance, distinctiveness could be measured by calculating the distances between crimes committed by different offenders, where higher scores would indicate higher levels of distinctiveness.

As an example of what behavioural distinctiveness might look like in the case of crime site selection, consider the map in Figure 1.2. This map represents all possible burglary targets in a particular geographic area as well as the spatial behaviour of two serial burglars, B1 and B2, who have each committed four crimes. Unlike the case in Figure 1.1, both burglars exhibit relatively low levels of similarity with respect to their crime site selection, which is indicated by the existence of long intra-offender distances. However, in this case, the two burglars do target different geographic areas compared to one another as indicated by the existence of relatively long inter-offender distances (an example of which is indicated by the solid arrow). It can therefore be said that the burglars do show high levels of distinctiveness with respect to their crime site selection. If the burglars in this case exhibited shorter inter-offender distances, they would be exhibiting lower levels of distinctiveness.

Figure 1.2. Behavioural distinctiveness (crime site selection)

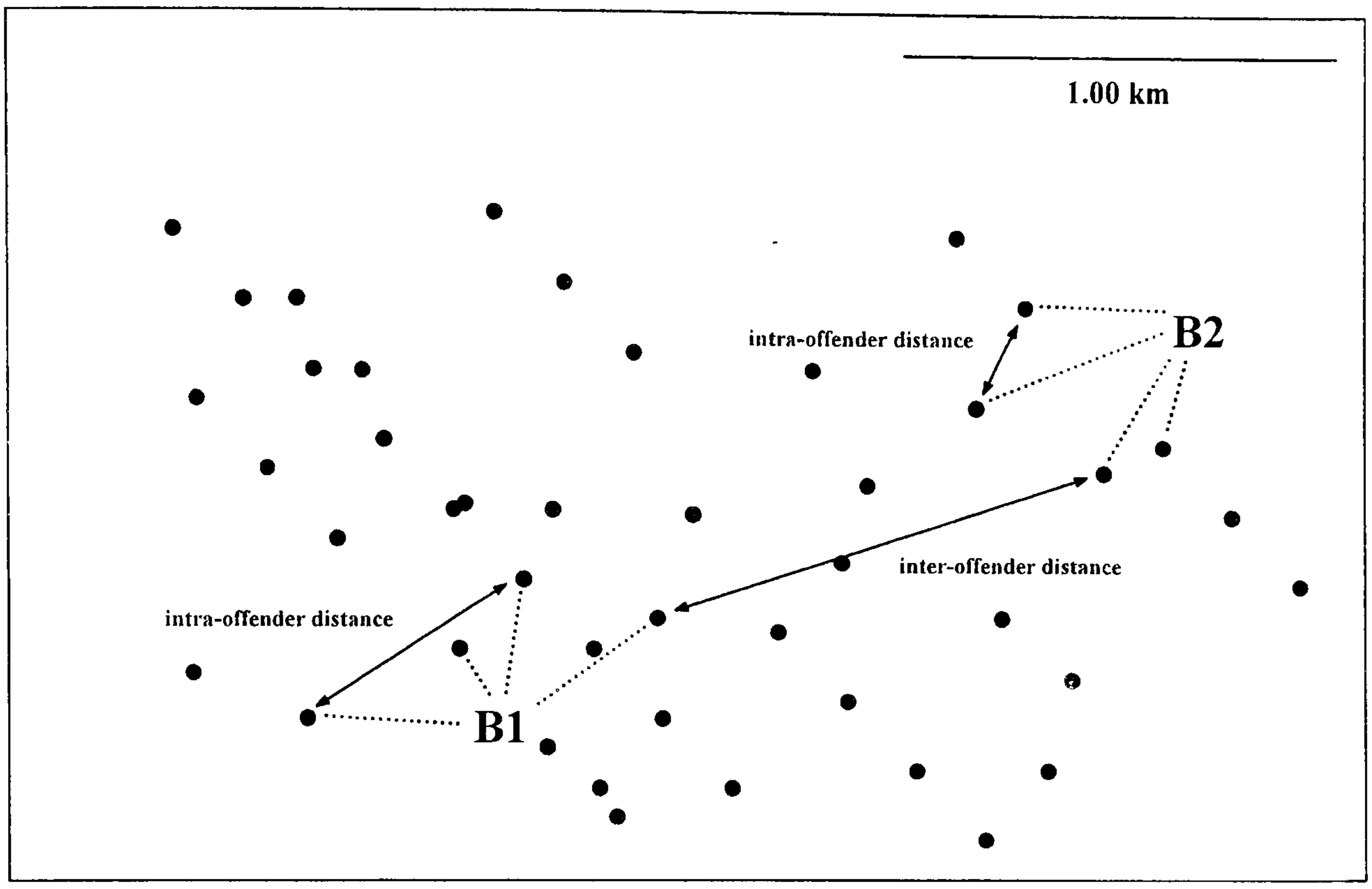


1.3.3. Behavioural consistency in serial burglary

Behavioural consistency can be defined as the extent to which burglars exhibit both similar and distinct behaviours across the burglaries they commit. This definition of consistency is very similar to the definition originally proposed by Canter (1995). He pointed out that, "...offender consistency consists of two components: the degree of variation within one offender's actions and the range of variation across a number of offenders" (p. 348). This definition is also similar to the way consistency has been defined in other contexts as well. For example, in the field of personality psychology, consistency has always been defined as the degree to which people express their individual differences in a stable fashion across situations (Bem & Allen, 1974; Funder & Colvin, 1991, Mischel & Peake, 1982; Shoda, 1999).

Based on this definition, consistency could be measured in a sample of burglars by comparing the similarity scores that result from an analysis of unlinked and linked serial burglaries. Consistency would be said to exist when there is a large discrepancy between the intra-offender and inter-offender similarity scores. As an example of what behavioural consistency might look like in the case of crime site selection consider the map in Figure 1.3. As was the case in Figure 1.1, both burglars exhibit high levels of similarity with respect to their crime site selection. However, as was the case in Figure 1.2, both burglars also exhibit high levels of behavioural distinctiveness with respect to their crime site selection. This combination of short intra-offender distances and long inter-offender distances is what makes B1 and B2 consistent. If the burglars in this case exhibited longer intra-offender distances or shorter inter-offender distances, they would be exhibiting lower levels of consistency.

Figure 1.3. Behavioural consistency (crime site selection)



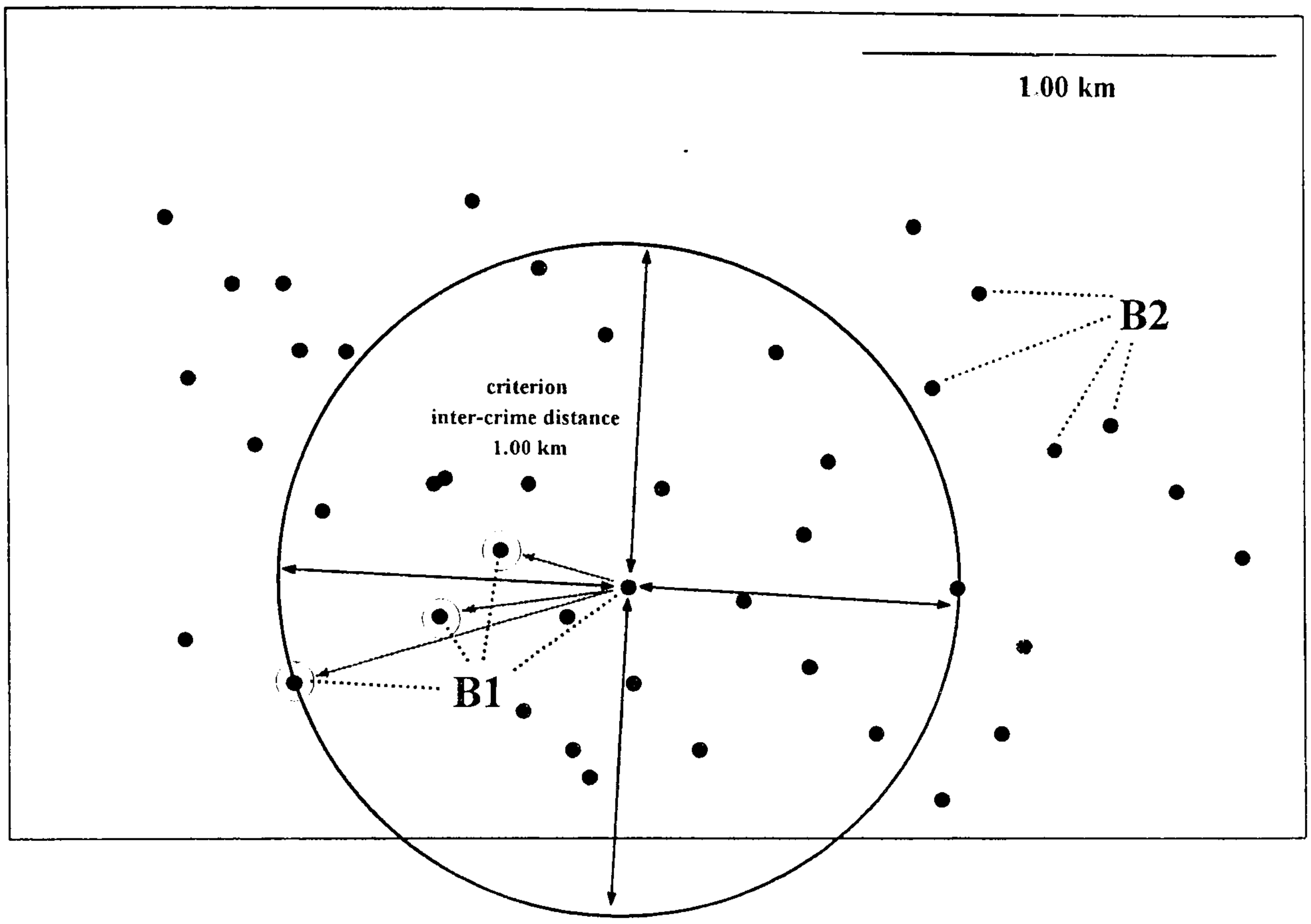
1.3.4. Behavioural discrimination in serial burglary

Behavioural discrimination can be defined as the extent to which burglaries committed by different burglars can be distinguished from one another. Based on the definitions proposed above, the extent to which this is possible depends on how consistent serial burglars are. In other words, discrimination would be possible if two or more burglars exhibit similar behaviours across their own crimes, the behaviours exhibited by each burglar are relatively distinct, and some decision criterion exists that accurately distinguishes between the offenders. In order to measure discrimination, decisions must be made about whether the same offender has committed a set of crimes, and these decisions must then be tested for accuracy by comparing them with the correct decisions.

As an example of what behavioural discrimination might look like in the case of crime site selection consider the map in Figure 1.4. As was the case in Figure 1.3, the burglars exhibit high levels of consistency with respect to their crime site selection. In addition to this consistency, however, a degree of across-crime similarity exists that can accurately discriminate between the two sets of crimes. In this case, a specific inter-crime distance can be selected, where all of the inter-crime distances that are shorter than it come from crimes committed by the same offender, and all of the inter-crime distances that are longer than it come from crimes committed by different offenders. This inter-crime distance can therefore be used as a criterion for making accurate linking decisions.

For example, when an inter-crime distance of 1.00 km is used as the decision criterion in Figure 1.4, so that all the crime pairs with an inter-crime distance shorter than this are treated as if they are linked, no incorrect decisions will be made. This is illustrated for just one of the crimes in Figure 1.4 – the burglary committed by B1 that is furthest to the right. It can be seen that the inter-crime distance of 1.00 km results in a 'linkage area' around this crime site with a diameter of 2.00 km. It can also be seen that the only burglaries occurring within this linkage area are the other crimes committed by B1 (indicated by double circles). As a result, perfect discrimination can occur – all linked crime pairs can be treated as linked and all unlinked crime pairs can be treated as unlinked. Indeed, for the serial burglaries represented in Figure 1.4, perfect discrimination accuracy can occur regardless of what crime site is used to develop the linkage area, so long as an inter-crime distance of 1.00 km is used.

Figure 1.4. Behavioural discrimination (crime site selection)



1.4. Measuring consistency and discrimination in the criminal context

It should be clear from the previous section that the behavioural processes examined throughout this research are related to one another. Based on the definitions just proposed, behavioural similarity and distinctiveness combine to produce behavioural consistency. In addition, the possibility of discriminating between crimes committed by different offenders depends on the extent to which the offenders exhibit consistency across their crimes. Thus, while each of the different processes have to be defined separately, in order to understand how they can be measured they need to be presented within an overall framework that shows how they relate to one another.

Different analytical procedures currently exist for measuring consistency and discrimination as they have just been defined. It will be argued in this chapter that the measurement of consistency in serial burglary behaviour directly corresponds to a traditional analysis of variance (ANOVA) problem, though it is not always thought of as such. Discrimination, on the other hand, can also be measured using a range of existing statistical procedures, although it requires the

ANOVA problem to be completely turned around. Indeed, the procedures that are usually employed to tackle these two different sorts of problems have often been thought of as exact opposites (Guttman, 1941, 1981, 1988).

To understand why this is the case, consider what is known about the two separate problems. When measuring behavioural consistency, we have two known populations, one consisting of intra-offender observations and another of inter-offender observations. The general goal is to determine whether differences exist across the two groups of observations with respect to some measure of across-crime similarity. When measuring discrimination, however, we are presented with some measure of across-crime similarity for a particular pair of crimes, but it is not known from which population the similarity score has come from. The goal here is to determine as accurately as possible whether the score comes from a population of intra-offender observations or inter-offender observations.

1.4.1. Analysing variance as a way of measuring consistency

The problem in ANOVA is whether significant differences exist across groups of observations (the independent variable) with respect to the mean of a particular score (the dependent variable). Using an ANOVA framework, the problem is solved by comparing two estimates of variance, one estimate coming from differences among scores within each group and another estimate coming from differences in the group means (Iversen & Norpoth, 1987). The test of whether or not the groups differ with respect to the particular variable is referred to as an F test. The test statistic, F, is computed as the ratio of between group variance over the pooled, or average, within group variance. If these two estimates are very similar, one can conclude that there is no significant difference between the groups. On the other hand, if the group means differ more than expected it can be concluded that there is a significant difference between the groups.

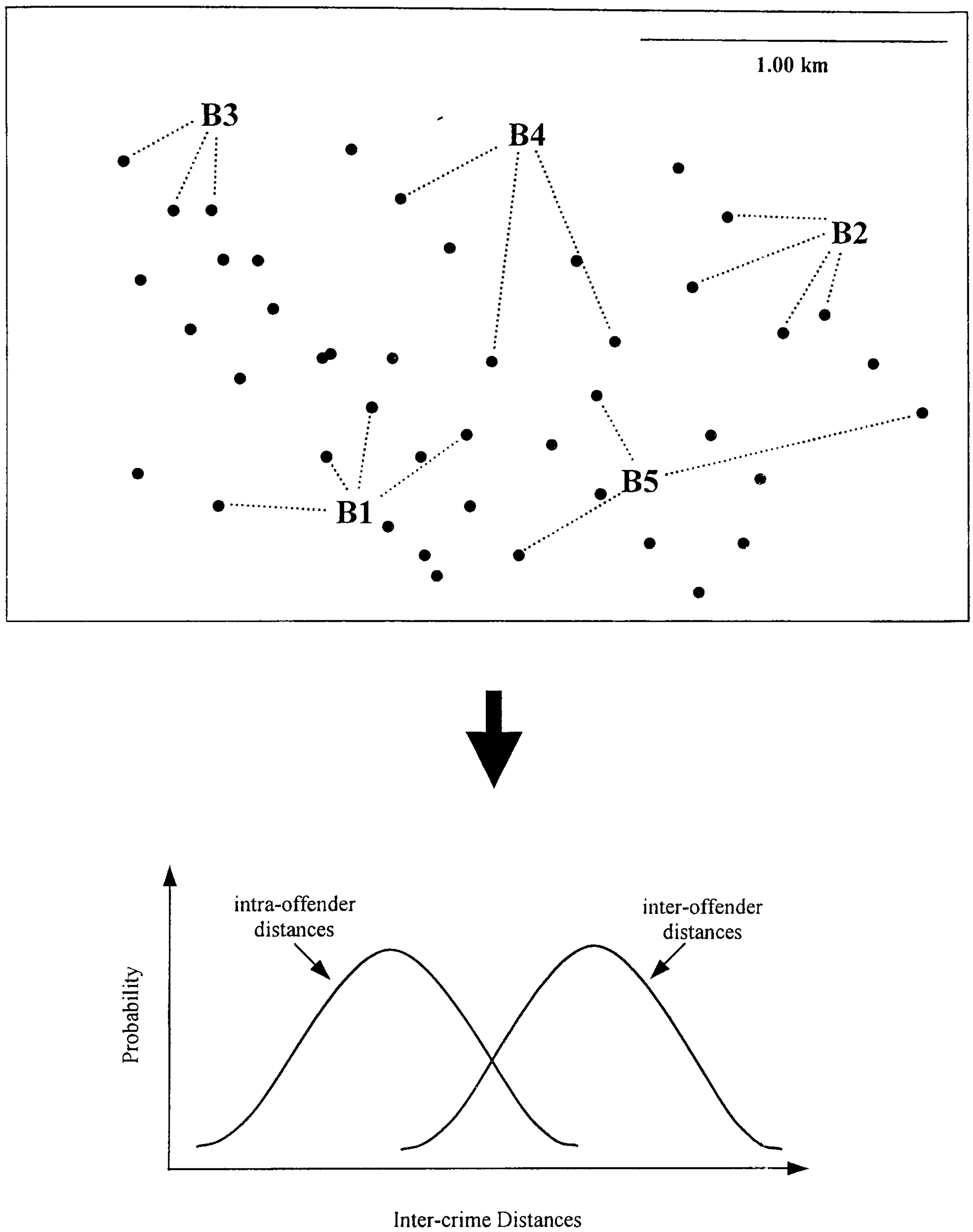
A special case of ANOVA arises when determining whether a difference exists across just two groups with respect to the mean of a particular variable. Typically, this problem is tackled using a t-test, but the t-value that results from such a test can simply be squared to become an F ratio as used in ANOVA

(Iversen & Norpoth, 1987; Scheffe, 1959). As indicated above, this applies to the present research because intra-offender and inter-offender observations of burglary behaviour are treated as two separate groups across which differences in behavioural variation are expected. In this case, if the level of behavioural variation is lower between crimes committed by different offenders than it is within crimes committed by the same offender, behavioural consistency does not exist. Stated in terms that are more in line with the proposed definitions, if the level of behavioural similarity is greater between crimes committed by different offenders than it is within crimes committed by the same offender, behavioural consistency does not exist.

Typically, the ANOVA problem is conceptualised using distributions of scores that are related to the variable of interest. This can easily be done in the present case as well by using the inter-crime distances examined throughout this chapter. For example, the map in Figure 1.5 represents all possible burglary targets in a particular geographic area as well as the spatial behaviour of five serial burglars. The graph below this map represents the distribution of inter-crime distances calculated between every pair of crimes. Specifically, the x-axis in this graph represents increasingly large inter-crime distances and the y-axis represents the probability that a pair of crimes will be associated with a particular distance. For the burglaries in Figure 1.5, these distributions indicate that the majority of burglars in the sample exhibit high levels of similarity and distinctiveness with respect to their crime site selection (i.e., short intra-offender distances and long inter-offender distances). As a result, these serial burglars are exhibiting high levels of consistency.

Such a finding would also be reflected by a relatively high F ratio in ANOVA terminology. This is because the level of inter-offender similarity is low (equivalent to high between group variance) and the level of intra-offender similarity is high (equivalent to low within group variance). This would not have been the case if inter-offender similarity were high (the burglars show no behavioural distinctiveness) or if intra-offender similarity were low (the burglars show no behavioural similarity). Thus, the greater the degree of overlap between the intra-offender and inter-offender distributions of scores, the lower the F ratio will become. The lower the F ratio becomes, the lower the level of behavioural consistency that is being expressed by a sample of serial burglars.

Figure 1.5. Constructing distributions of inter-crime distances



1.4.2. Rephrasing the ANOVA problem as a discrimination problem

The problem of discriminating between crimes committed by different offenders is the inverse of the consistency problem. With the consistency problem, the goal is to determine whether group membership produces a significant difference on some sort of similarity score. With the discrimination problem, the goal is to determine whether that similarity score can be used to predict group membership. With the consistency problem, the independent variable refers to the intra-offender and inter-offender observations and the dependent variable refers to the across-crime similarity score. With the discrimination problem, the independent variable refers to the across-crime similarity score and the dependent variable refers to the intra-offender and inter-offender observations.

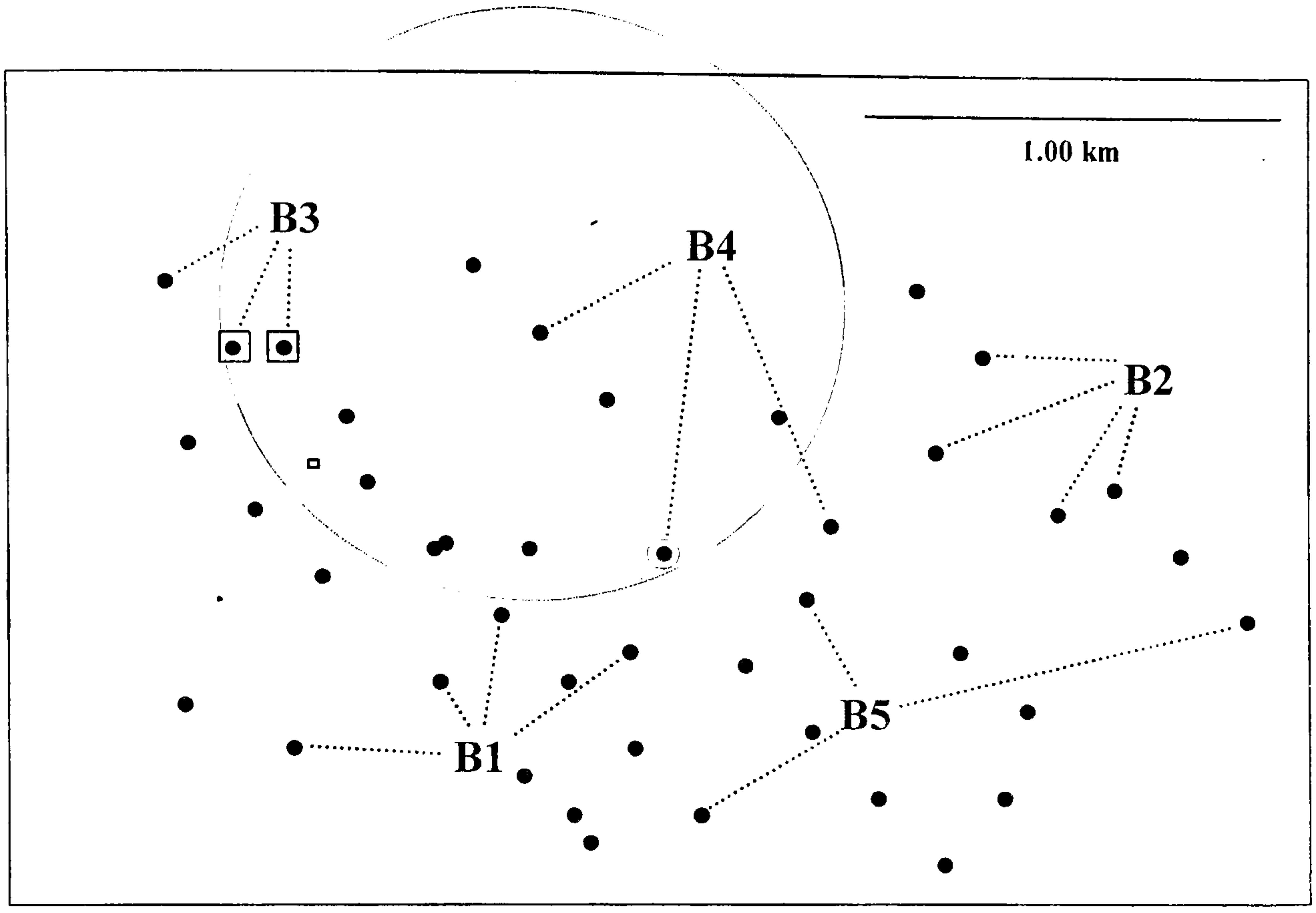
Despite these differences, the consistency and discrimination problem are fundamentally related. As the definition of behavioural discrimination proposed above makes clear, the level of discrimination that is possible in any given sample of serial burglars depends on how consistent those serial burglars are. That is, the higher the level of consistency exhibited by the burglars, the higher the level of discrimination that will be possible as a result. In other words, if group membership does produce a significant difference with respect to some sort of across-crime similarity score, it is also likely that this across-crime similarity score will be able to discriminate between the groups (Guttman, 1988).

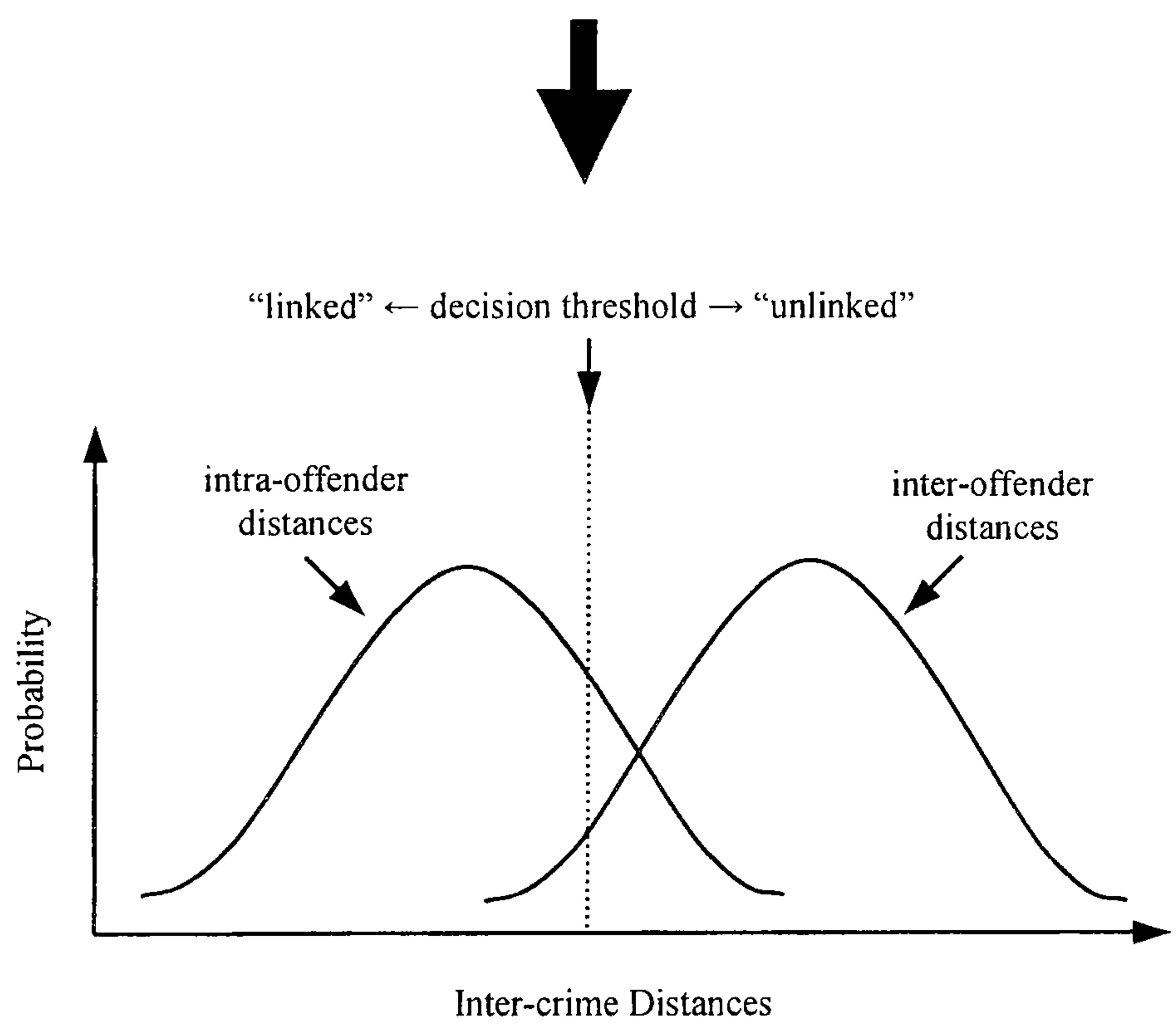
There is a wide range of procedures that are currently available for trying to solve the discrimination problem (e.g., discriminant function analysis). Although the procedures sometimes work in different ways, the primary goal of each is to find a set of variables that maximise the separation between different groups of observations. Often, these procedures can be used in conjunction with techniques such as ANOVA. Where the procedures designed to tackle the two separate problems really start to differ is when it comes to the process of classification. Classification entails the use of a variable, or sometimes a set of variables, to assign people or cases to particular groups (Tabachnik & Fidell, 1996). The goal is to classify these people or cases as accurately as possible, which is tested by comparing actual group membership to predicted group membership. From this, it is possible to calculate the number of cases correctly and incorrectly classified as well as the nature of these classifications.

In order to classify two burglaries as being either linked or unlinked, a decision threshold of some kind needs to be established. Swets (1973) refers to a decision threshold as a cut-off point along a continuum of evidence whereby any value obtained above that point (or below it, depending on the nature of the evidence) results in a positive decision. In the present research, this decision threshold will always correspond to an across-crime similarity score of some kind, and the positive decision will always be that crimes are linked. An example of this procedure was already illustrated in Figure 1.4 where an inter-crime distance of 1.00 km was used as a threshold for deciding when two crimes were linked. However, decision thresholds are more often conceptualised in a graph like the one in Figure 1.5.

This can easily be done in the present case by using an inter-crime as a decision threshold. For example, the map in Figure 1.6 once again represents all possible burglary targets in a particular geographic area as well as the spatial behaviour of five serial burglars. The graph below this map represents the distribution of inter-crime distances calculated between every pair of crimes, which suggest that the serial burglars are exhibiting high levels of consistency with respect to their crime site selection. In this graph, however, a decision threshold has also been set along the x-axis, which indicates when a positive decision should be made. In this case, any time a pair of crimes is less than 1.00 km apart, those crimes will be treated as linked. This is illustrated for one of the crimes in Figure 1.6 – the burglary committed by B4 that is furthest to the left. In this case, one of the crimes that would be linked to this crime is another crime committed by B4 (indicated by a double circle). However, two crimes committed by B3 would also be incorrectly linked (indicated by boxed circles).

Figure 1.6. Assigning a decision threshold





The result of setting such a decision threshold is that the degree of discrimination that is possible in a sample of serial burglars can now be measured by examining the accuracy of the resulting classification decisions. What the graph in Figure 1.6 makes clear is that, in this particular case, the decision threshold adopted results in many more correct classifications than incorrect classifications. As a result, behavioural discrimination would be considered possible. Note, however, that this would not have been the case in Figure 1.6 if the intra-offender and inter-offender distributions of inter-crime distances had overlapped to a greater degree.

1.5. What this means for the development of an analytical procedure

Determining the extent to which offenders display similarity and distinctiveness is a straightforward task because levels of intra-offender or inter-offender similarity simply need to be examined. However, developing a procedure that can determine the extent to which consistency exists is far more difficult, as is developing a procedure that can determine the extent to which discrimination is possible.

While procedures currently exist for tackling these problems, they are cumbersome and are unlikely to be of great use in the investigative context. A more effective approach would be to develop a procedure that is capable of (1) analysing consistency and discrimination simultaneously and (2) presenting the results in a way that is of use to investigators in the field. A potential procedure for accomplishing this is proposed in the next chapter. However, some of the key features that such a procedure would need to take into account are briefly discussed here.

1.5.1. Distribution overlap as a measure of consistency

It has been shown that it may be possible to measure the level of consistency exhibited by a sample of serial burglars by examining the degree of overlap that exists between intra-offender and inter-offender distributions of similarity scores. As was illustrated, if burglars display relatively high levels of similarity across their own crimes, and the behaviours they exhibit are relatively distinct, then the intra-offender and inter-offender distributions of similarity scores that result from an examination of their behaviour will overlap very little. Therefore, for any analytical procedure to be useful, it would have to be able to quantity this degree of overlap.

1.5.2. Classification accuracy as a measure of discrimination

It has also been shown that it may be possible to measure the level of discrimination possible in a sample of burglaries by examining the accuracy of classification decisions that result when using a decision threshold. If the use of a decision threshold results in more correct decisions than incorrect decisions, some degree of discrimination can be said to exist. However, it is clear from the previous discussion of the problem that three additional challenges must be overcome in order to use this method as a basis for measuring discrimination. The first challenge emerges because of the relationship between consistency and discrimination. The second challenge emerges because of the fact that decision thresholds can be set anywhere along a continuum of similarity scores. The third challenge emerges because numerous decision outcomes are possible when discriminating between crimes committed by different offenders.

(a) The relationship between consistency and discrimination

As already pointed out, based on the way that consistency and discrimination have been defined here, there is a direct relationship between the two processes. Essentially, as the level of consistency increases in a given sample of serial burglars so to will the overall level of discrimination that is possible. On the one hand, if intra-offender and inter-offender distributions do not overlap at all it would be possible to achieve perfect discrimination. On the other hand, if intra-offender and inter-offender distributions overlap completely it would be difficult to achieve levels of discrimination that are greater than chance. Therefore, for any analytical procedure to be useful, this relationship between consistency and discrimination will need to be appreciated.

(b) Accounting for multiple decision thresholds

Decision thresholds can be placed anywhere along a continuum of similarity scores, not just at a single point as in Figure 1.6, and the exact position of the threshold will effect the level of discrimination that is possible. Even for a set of intra-offender and inter-offender distributions that overlap very little, discrimination performance can potentially range from very accurate to very inaccurate. Therefore, for any analytical procedure to be useful, it must be able to evaluate the extent to which discrimination accuracy changes as the decision threshold is varied. Ideally, an effective analytical procedure would also be capable of specifying what an optimal decision threshold is.

(c) Measuring all potential decision outcomes

When discriminating between crimes committed by different offenders there are a number of possible outcomes. At the most basic level, correct and incorrect decisions can be made and discrimination accuracy can be calculated by considering these decisions. However, in reality, there are two types of correct decisions that can be made and two types of incorrect decisions (see Table 1.1). For example, given two crimes committed by the same offender, one can decide that the pair is a linked crime pair or an unlinked crime pair. These types of decision outcomes are often referred to as hits and misses respectively (Swets, 1996). On the other hand, given two crimes committed by different offenders, one can also decide that the pair is a linked crime pair or an unlinked crime pair. These types of decision outcomes are often referred to as false alarms and correct rejections (Swets , 1996).

Table 1.1. Possible decision outcomes in the discrimination task

		Reality:	
		Linked	Unlinked
Prediction:	Linked	hit	false alarm
	Unlinked	miss	correct rejection

When carrying out the discrimination task, providing an indication of the number of correct decisions that are made without also providing an indication of how many incorrect decisions are made would provide a very biased view of discrimination accuracy. However, the different types of correct and incorrect decisions also need to be accounted for. This is because the benefits associated with the different types of correct decisions and the costs associated with the different types of incorrect decisions may differ substantially. As just one example, deciding that a crime pair is unlinked when it is in fact linked may be considered much more costly in residential burglary compared to commercial burglary because of the high degree of personal suffering that residential burglary victims experience (Brown & Harris, 1989; Maguire, 1980). Therefore, for any analytical procedure to be useful, it must be able to evaluate all possible decision outcomes that can result when attempting to discriminate between crimes committed by different offenders.

CHAPTER 2
A PROCEDURE FOR EXAMINING CONSISTENCY AND DISCRIMINATION

2.1. Introduction

As described in Chapter 1, any analytical procedure developed for the purpose of examining behavioural consistency and discrimination in serial burglary should satisfy a number of conditions. For example, it should be apparent from the output of an effective procedure how consistently burglary behaviour is being expressed, regardless of what aspect of behaviour is explored or how that behaviour is measured. It will be argued in this chapter that one possible way of doing this is to develop a procedure that has the capability to quantify the extent to which intra-offender and inter-offender distributions of across-crime similarity scores overlap.

An effective procedure must also provide some systematic way of deciding how similar two crimes must be before it should be decided that the same offender has committed them, and it should give some indication as to the consequences of making these sorts of decisions using different levels of similarity. It will be argued in this chapter that one possible way of doing this is to develop a procedure that has the capability to quantify the probability of making all possible linking decisions, both correct and incorrect, at each and every level of across-crime similarity.

The few procedures that exist at present for examining behavioural consistency and discrimination in the criminal context fall short in their coverage of these important aspects. There is a need, therefore, to develop a more suitable analytical procedure. Within this chapter, receiver operating characteristic (ROC) analysis will be proposed as one possible procedure that can satisfy all the conditions outlined above. It will be demonstrated that this method of analysis relates directly to the behavioural processes discussed in Chapter 1. As a result, it is an extremely useful framework for examining consistency and discrimination in serial burglary.

2.2. The origins of ROC analysis

Before discussing the specific details of the proposed analytical procedure, it may be useful to put it in historical context. ROC analysis has its roots in psychophysics, statistical decision theory and signal detection theory (Swets, 1973). In addition, it is now a common procedure being used to examine a wide range of sensory and cognitive processes (Green & Swets, 1974; Rose, 1995), and to evaluate decision-making performance in diagnostic fields (Swets, 1996; Swets & Pickett, 1982; Swets, Dawes & Monahan, 2000a, 2000b).

The emphasis in the first part of this chapter will be on providing a general overview of some key components of psychophysics, statistical decision theory, signal detection theory, and diagnostic decision-making, as they provide the basis for understanding the ROC procedure in the present context. The second part of this chapter will provide a more formal presentation of the procedure, to show how it can be used as a method for examining behavioural consistency and discrimination in serial burglary.

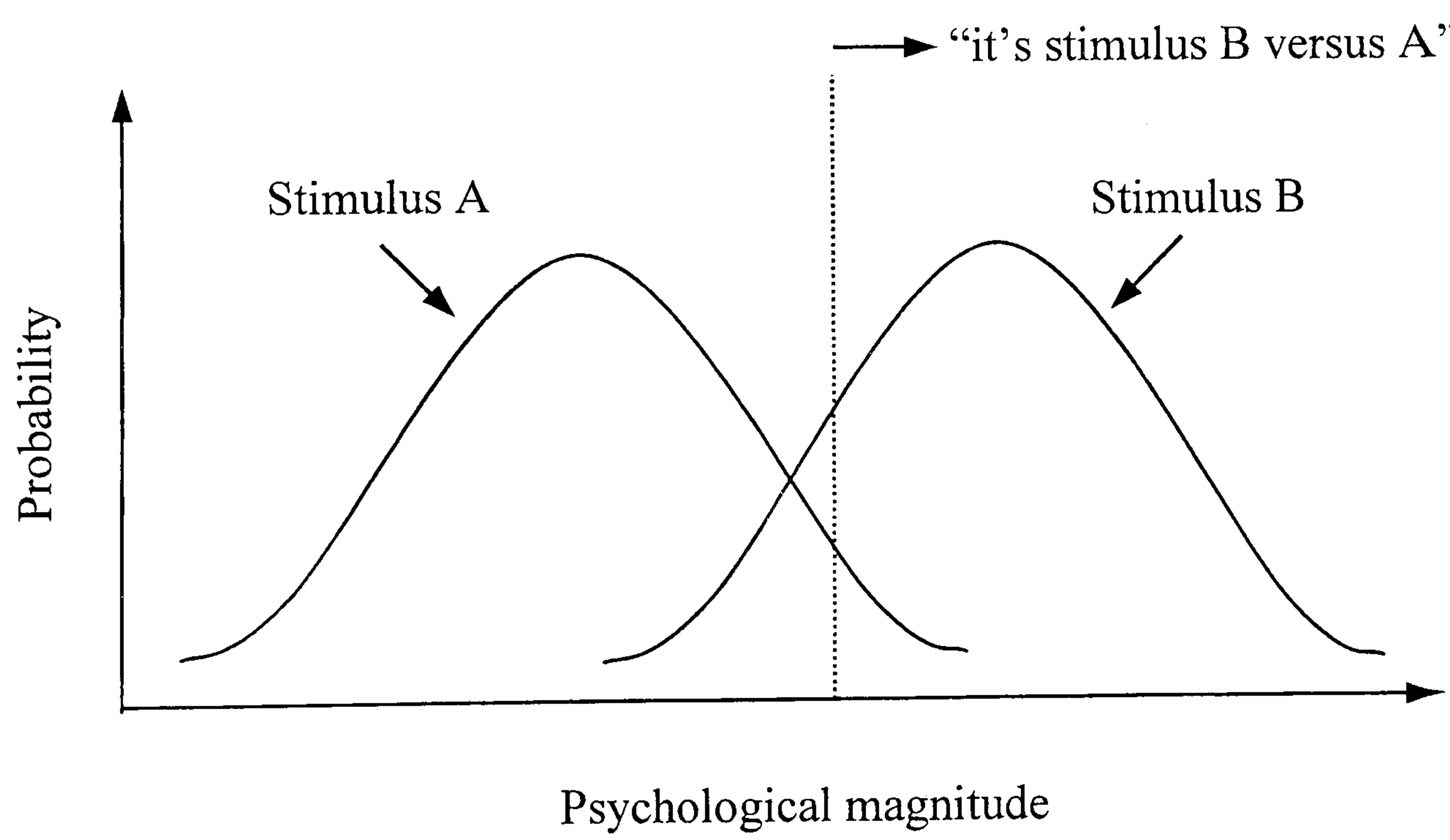
2.3. Psychophysics

The ideas behind ROC analysis have their original roots in studies carried out by psychophysicists such as Fechner, Thurstone and Blackwell (Swets, 1973). Fechner (1860/1966) was one of the first people of tackle the issue of discrimination when he attempted to determine the relationship between the various attributes of sensation (e.g., intensity) and the physical measurements of stimuli. Primarily, Fechner was interested in two issues: (1) at what point can people discriminate one stimulus from another (the so-called just noticeable difference), and (2) at what point can people detect the presence of a signal (the so-called stimulus just noticeable). Emerging from this research was Fechner's now famous psychometric function, an indication of how the proportion of positive responses (e.g., "it's stimulus A versus B") relates to various measures of stimulus strength or stimulus difference.

In contrast to Fechner, Thurstone (1927) focused almost exclusively on the recognition problem (i.e., discriminating one stimulus from another). According to Swets (1973), Thurstone expanded on the ideas proposed by Fechner by assuming the stimuli to be distinguished could be represented as overlapping probability distributions along some scale of psychological or sensory magnitude (see Figure 2.1). Specifically, Thurstone assumed an observer would select stimulus B instead of stimulus A whenever the magnitude of stimulus B exceeded the magnitude of stimulus A.

Blackwell (1952) continued to use the idea of overlapping probability distributions in his thinking, though he was more concerned with the detection task (i.e., detecting the presence of stimulus A). In addition, Blackwell assumed the existence of a decision criterion or threshold for a positive response, which would virtually eliminate the possibility of mistaking stimulus A for stimulus B (i.e., the dashed line in Figure 2.1). By assuming this, however, it meant that sensory magnitudes falling below this criterion would be indistinguishable, implying a physiological threshold within people whereby stimuli presented below this threshold would result in confusion.

Figure 2.1. Hypothetical distributions representing two stimuli

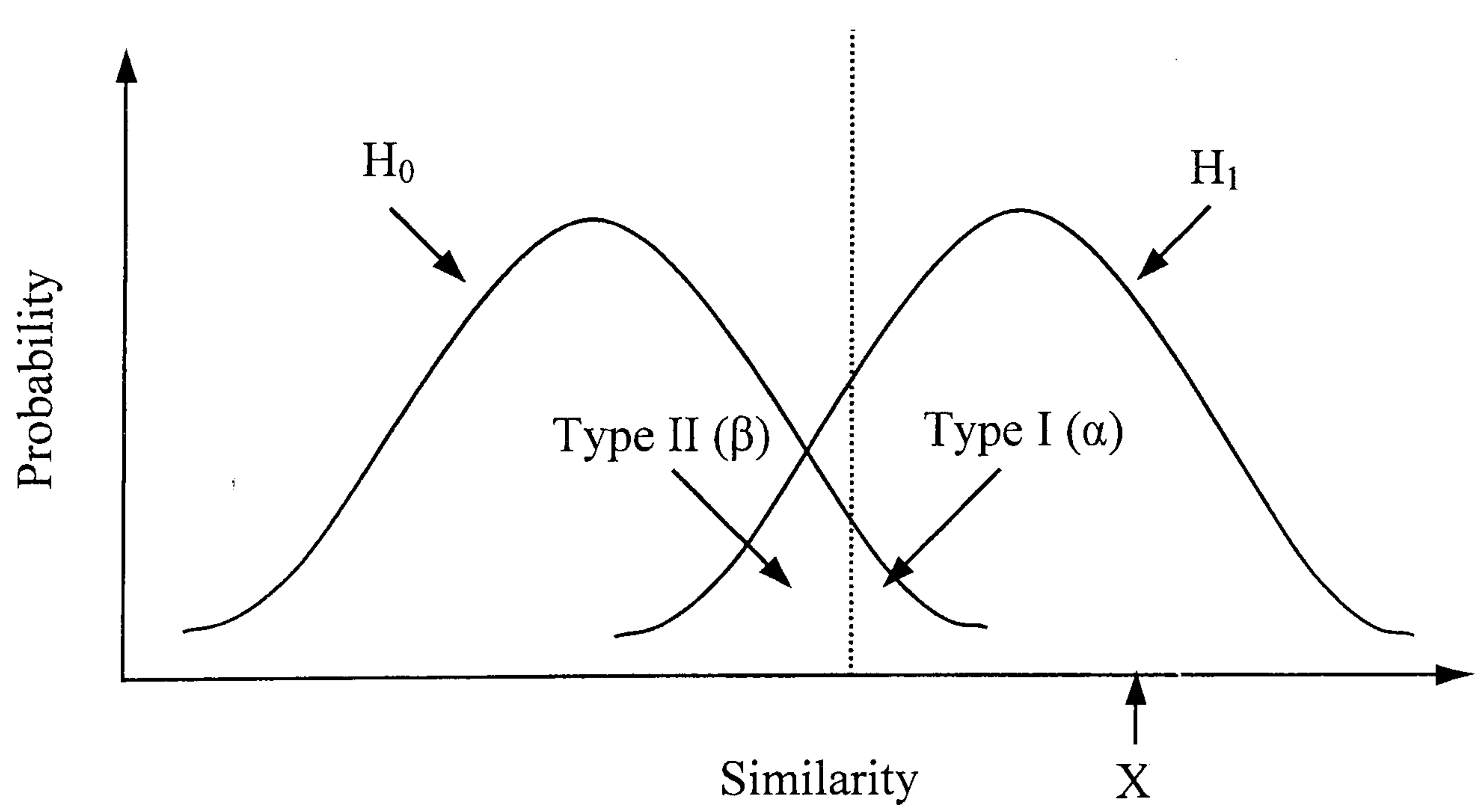


2.4. Statistical decision theory

Hypotheses, as defined in statistical decision theory, are often represented as probability distributions similar to those presented above. A pair of crimes may exist, for example, with a similarity score equal to X and one may wish to test the hypothesis that such a score came from a population of crime pairs committed by different offenders. We could set up a null hypothesis, H_0, which states that the similarity score does come from the population of inter-offender crimes (the left distribution in Figure 2.2). We could also set up an alternative hypothesis, H_1, which states that the similarity score does not come from the population of inter-offender crimes, it is in fact higher and therefore represents an intra-offender crime pair (the right distribution in Figure 2.2).

If we knew the mean and standard deviation of inter-offender similarity scores, we could then determine if the similarity score, X, was high enough to indicate it was unlikely to come from the population of inter-offender similarity scores. If this were found to be the case, we could reject the null hypothesis and decide that the similarity score came from observations of crimes committed by the same offender.

Figure 2.2. Hypothetical distributions representing two hypotheses



2.4.1. The decision threshold in statistical hypothesis testing

Deciding what 'high enough' means in the above example is equivalent to setting a decision threshold somewhere along the x-axis in Figure 2.2, much like the decision criterion talked about by Blackwell. This decision threshold, more commonly referred to as the significance level or rejection level in statistics, determines how high the similarity score in the previous example must be in order to decide it does not come from the population of inter-offender similarity scores. Scores below the threshold would lead to the acceptance of H_0 (more precisely, a failure to reject H_0) while scores above it would lead to the acceptance of H_1.

In fact, where this threshold is set determines the relative probabilities of all possible decision outcomes in situations where two competing statistical hypotheses are being tested. In such situations, there is always a chance of making two types of correct decisions: (1) rejecting a null hypothesis that is really false and (2) failing to reject a null hypothesis that is really true. However, there is also a chance of making two types of incorrect decisions: (1) rejecting a null hypothesis that is really true (a Type I error) and (2) failing to reject a null hypothesis that is really false (a Type II error).

Given a decision threshold set at a particular position along the x-axis, each of these possible decisions occurs with a certain frequency. These frequencies are represented in Table 2.1 by the letters *a*, *b*, *c* and *d*. For example, the letter *a* refers to the number of times H_0 is rejected when it is in fact false, for a particular decision threshold or significance level. These raw frequencies can be converted into conditional probabilities, represented by the values p in Table 2.1. These are estimated probabilities of particular decisions being made conditional upon particular realities.

Specifically, the conditional probability of making a Type I error, designated α, is equal to the proportion of the area under the H_0 distribution to the right of the decision threshold in Figure 2.2. This probability can be calculated by dividing a by $a+c$. Since there is only one other possible decision that can result when H_0 is true, the conditional probability of failing to reject a null hypothesis that is true is the complement of α, or $1-\alpha$, calculated by dividing c by $a+c$. The conditional probability of making a Type II error, designated β, is equal to the proportion of the area under the H_1 distribution to the left of the decision threshold in Figure

2.2. This probability can be calculated by dividing b by $b+d$. Since there is only one other possible decision that can result when H_0 is false, the conditional probability of rejecting a null hypothesis that is really false is the complement of β, or $1-\beta$, calculated by dividing d by $b+d$. See Appendix A for a more thorough discussion of these conditional probabilities.

Table 2.1. Possible decision outcomes in statistical hypothesis testing

		Reality: H_0 false	Reality: H_0 true	
Decision:	Reject H_0	a Correct decision $p=1-\beta=a/(a+c)$	b Type I error $p=\alpha=b/(b+d)$	$a+b$
	Fail to reject H_0	c Type II error $p=\beta=c/(a+c)$	d Correct decision $p=1-\alpha=d/(b+d)$	$c+d$
		$a+c$	$b+d$	$a+b+c+d=N$

2.4.2. The relationship between Type I and Type II errors

An important aspect of statistical hypothesis testing is that it is possible to specify the probabilities of making the two types of errors, with the aim of course being to make them both zero. However, as Welkowitz, Ewen and Cohen (1982) point out, this is not possible because the two probabilities are fundamentally related in such a way that decreasing one always makes the other more likely, and vice versa.

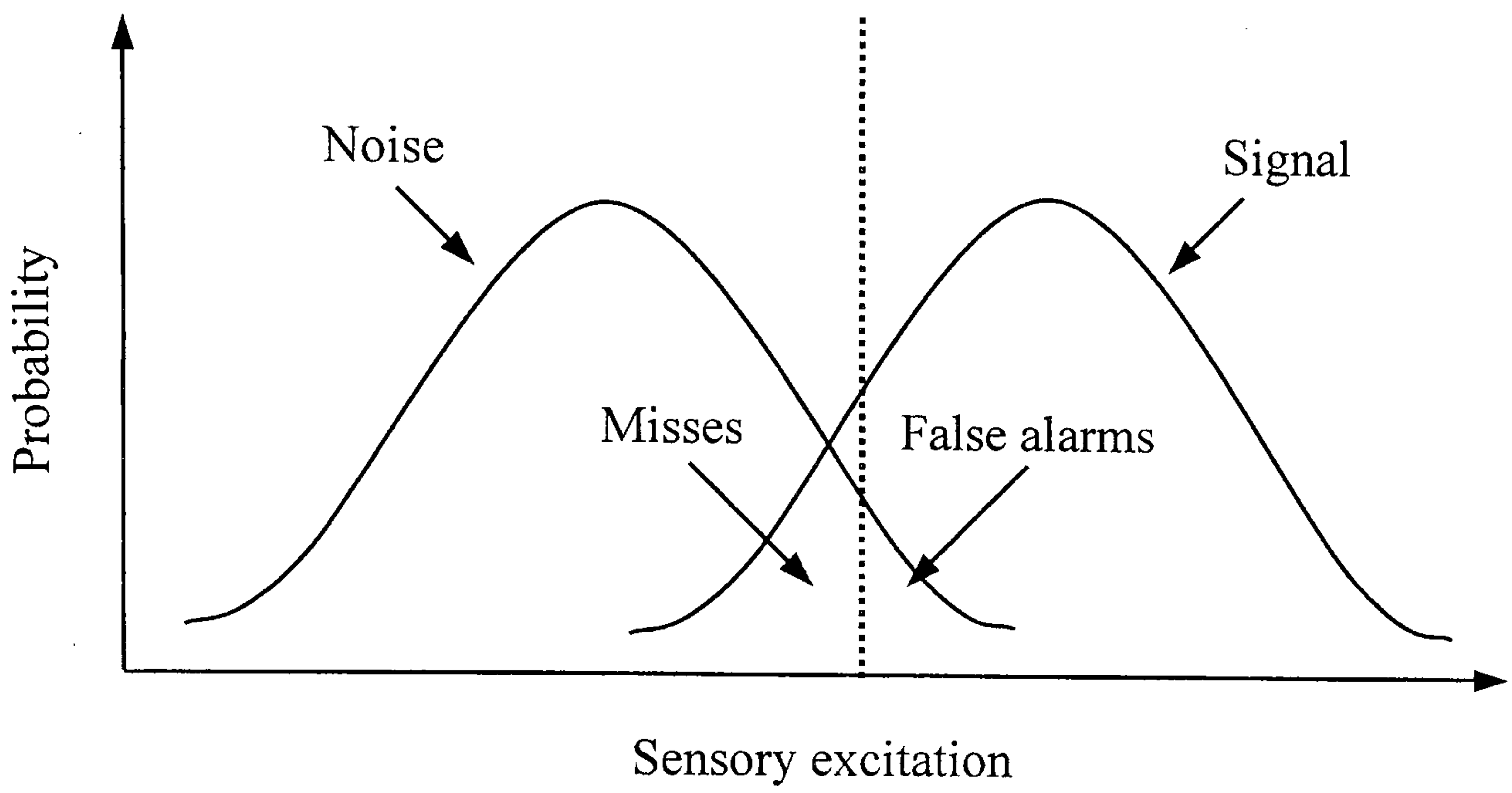
It can be seen in Figure 2.2, for example, that if the decision threshold were moved to the right, the area under the H_0 distribution to the right of the threshold would decrease (i.e., a Type I error would become less likely). By doing this, however, the area under the H_1 distribution to the left of the threshold increases (i.e., a Type II error would become more likely). Conversely, if the decision threshold were moved to the left in order to decrease the probability of making a Type II error, the probability of making a Type I error would simultaneously increase.

The choice of where to set the decision threshold when testing statistical hypotheses is a subjective one (Gigerenzer & Murray, 1987). However, various rules of thumb have been developed and are regularly used. The most common rule is to fix the probability of making a Type I error at 0.05 so that the chance of rejecting H_0 when it is actually true is very small (Neyman & Pearson, 1933). Such a rule encourages a degree of caution when testing statistical hypotheses so that new theories are accepted only when researchers can be confident they are correct. As a result, however, there is a risk of incorrectly rejecting theories that are in fact useful. In light of the fact that the probability of making a Type II error is often totally ignored when conducting statistical tests, this is a risk most scientists seem willing to take.

2.5. Signal detection theory

Developed in the early 1950's, signal detection theory was originally concerned with the detection of electronic radar signals (Peterson, Birdshall & Fox, 1954). In particular, researchers were interested in how accurately radar receivers could detect radar signals in the presence of background interference, or noise. Signal detection theorists viewed this task as a problem involving hypothesis testing (Swets, 1973). As Swets discusses, background noise was typically treated as a null hypothesis (the left distribution in Figure 2.3), while the presence of a signal was treated as an alternative hypothesis (the right distribution in Figure 2.3). The x-axis in the signal detection context was assumed to represent increasing levels of sensory excitation in the radar receiver, and the task, just as it is in all statistical hypothesis testing, was to decide which hypothesis the available evidence favoured (Swets, 1973).

Figure 2.3. Hypothetical distributions representing noise and signal



2.5.1. Hits, misses, correct rejections and false alarms

In signal detection terms, the Type I errors and Type II errors made when testing statistical hypotheses are referred to as false alarms and misses respectively, whereas the complements of these errors are referred to as hits and correct rejections. In the radar context, a false alarm would occur when a radar signal is not present but the radar receiver decides that one is. A miss occurs when a radar signal is present but the radar receiver decides that one is not. A hit occurs when a radar signal is present and the radar receiver decides that one is. Lastly, a correct rejection occurs when a radar signal is not present and the radar receiver decides that one is not. The conditional probabilities of each of these outcomes can be estimated as they were when testing statistical hypotheses. Furthermore, as in statistical hypothesis testing, the probability of making one type of error cannot be varied without having an effect on the other.

2.5.2. Discrimination and decision processes

One of the major advances made by signal detection theorists came from their recognition that the probability of these decision outcomes is not purely a function of the observers ability in a physiological sense to discriminate between signal and noise, as previously thought by many psychophysicists (Swets, 1973). Instead, the signal detection task consists of a decision process as well, where the observer sets a decision threshold and uses this as the basis for indicating whether a signal is present.

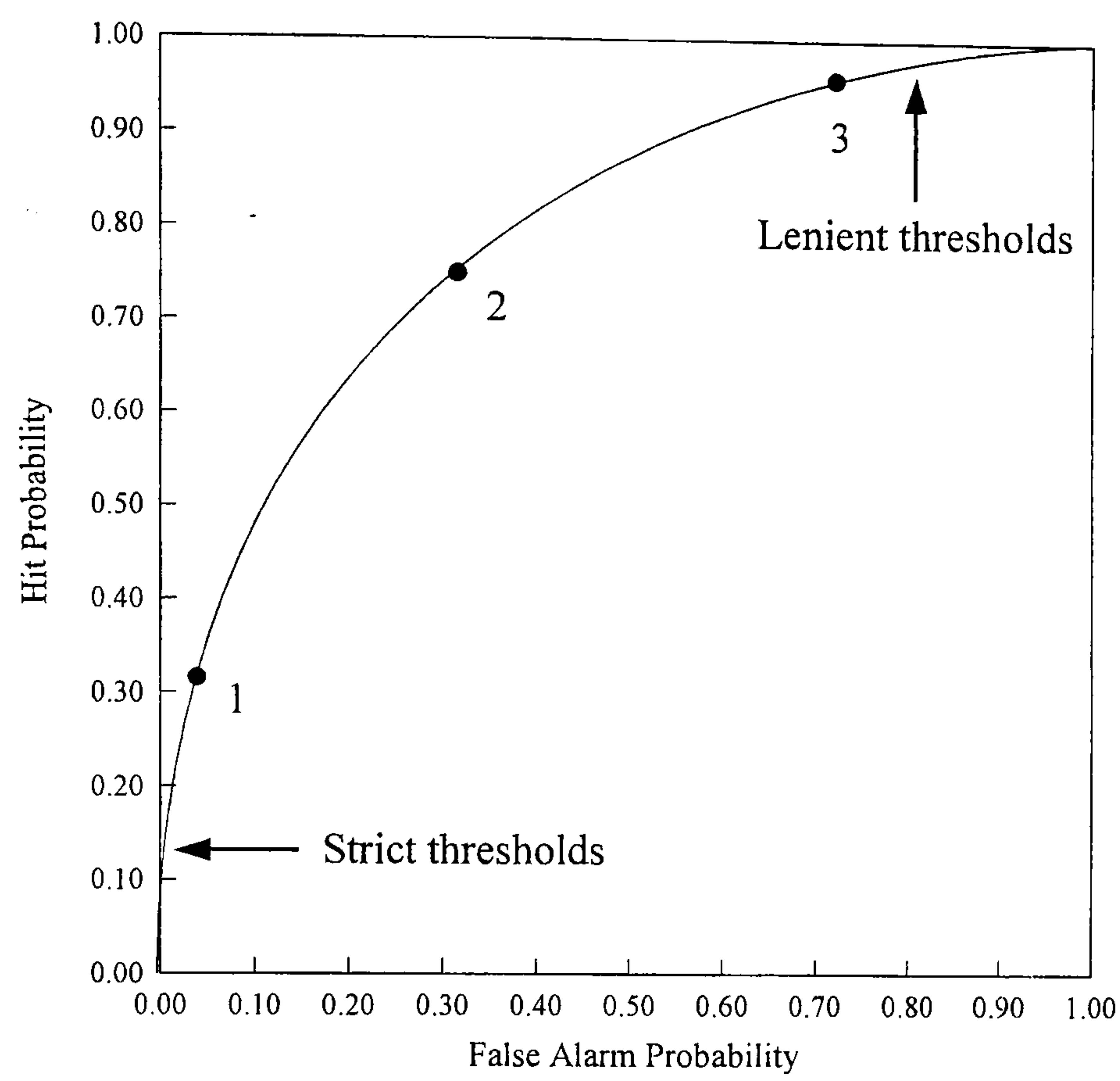
The particular decision threshold an observer chooses has been found to relate to a variety of factors, such as the perceived benefit of picking one alternative over the other or judgements about how often a signal will be presented (Swets, 1964). By altering these factors, signal detection theorists were able to show that subjects could reliably detect signals that were seemingly undetectable before. As a result, it became generally accepted that to obtain useful measures of discrimination accuracy this variable decision threshold had to be taken into account.

2.5.3. The ROC graph

The realisation of this fact led signal detection theorists to develop ROC analysis. ROC analysis differs from conventional statistical tests in at least two important ways. First, the emphasis in ROC analysis is not just on false alarms. Instead, all types of decisions (hits, misses, correct rejections and false alarms) are given equal importance. Second, the probabilities of making these decisions are not based on a single decision threshold. Instead, the probabilities are calculated across many different thresholds.

Very lenient decision thresholds are located towards the far left of the x-axis in Figure 2.3 on the previous page. Very strict decision thresholds are located towards the far right. ROC analysis essentially consists of plotting on a graph the hit probability, *p*H, against the false alarm probability, *p*FA, for decision thresholds that range from very lenient to very strict (see Figure 2.4). However, because the miss probability, *p*M, and the correct rejection probability, *p*CR, are the complements of these two probabilities, information about all four decision outcomes is contained on the graph.

Figure 2.4. A hypothetical ROC graph of pH versus pFA



For a single pair of signal-noise distributions, such as the distributions in Figure 2.3, plotting pH and pFA across a range of decision thresholds results in a concave downward curve starting in the lower left-hand corner of the plot and ending in the upper right-hand corner. Specifically, when a lot of evidence is required to decide that a signal is present, pH and pFA will both be low and result in ROC points falling on the lower end of the curve (e.g. threshold 1). On the other hand, when little evidence is required to decide that a signal is present, pH and pFA will both be high and result in ROC points falling on the upper end of the curve (e.g. threshold 3). Between these two extremes, moderate decision thresholds will result in ROC points falling along the remainder of the curve (e.g. threshold 2).

In order to obtain a different ROC curve, the properties of the signal simply need to be altered, making it easier or harder for the observer to detect. The intensity of the signal could be increased, for example, making the signal produce greater sensory excitation. This, in turn, would create less overlap between the signal and noise distributions in Figure 2.3, which would make distinguishing the signal from the noise less difficult. In this case, *p*H and *p*FA will still vary as a function of where the decision threshold is placed, but for any given value of *p*FA in the previous condition, *p*H is likely to be higher. The end result will be a ROC curve higher in elevation than the one in Figure 2.4.

If instead of increasing the intensity of the signal it were substantially decreased, the effect would be the opposite. In this case, the signal would produce less sensory excitation, creating greater overlap between the signal and noise distributions in Figure 2.3, which would make distinguishing the signal from the noise more difficult. When the new values of *p*H and *p*FA are plotted on a ROC graph, the result would be a ROC curve lower in elevation than the one in Figure 2.4.

2.5.4. An appropriate measure of discrimination accuracy

As can be seen in the above example, the degree of overlap between signal-noise distributions, and the level of discrimination that is ultimately possible as a result, is indicated by the height of the ROC curve (McNicol, 1972). Consequently, the usual method for measuring discrimination accuracy in signal detection tasks is to calculate the proportion of the graph's area lying beneath the curve, referred to as the area under the curve (AUC). As Swets (1973) points out, because this index specifies the locus of the entire ROC curve, rather than any single point along it, the AUC reflects all possible decision thresholds and hence is independent of any one. Thus, the AUC provides an unbiased indication of overall discrimination accuracy.

Signal-noise distributions that overlap completely typically lead to signal detection at a chance level of accuracy and result in ROC curves that fall along the positive diagonal. These ROC curves cut the ROC graph in half and therefore receive an AUC measure of 0.50. On the other hand, signal-noise distributions that do not overlap at all typically lead to perfect signal detection and result in ROC curves falling along the left and upper axes of the ROC graph. These ROC curves have the entire area of the ROC graph below them and therefore receive an AUC measure of 1.00. Although such judgements depend to some extent on contextual factors, Swets (1988) has arbitrarily set a guideline for determining how much accuracy an AUC reflects. He argues that AUCs of 0.50 are non-informative, AUCs between 0.50 and 0.70 indicate low levels of accuracy, AUCs between 0.70 and 0.90 indicate moderate levels of accuracy, and AUCs between 0.90 and 1.00 indicate high levels of accuracy[1].

2.5.5. Calculating the AUC

In order to calculate the AUC a variety of procedures can be used, the appropriateness of each depending on whether certain distributional assumptions are met. Without going into the details of each procedure, suffice it to say that a suitable AUC measure exists for signal-noise distributions that are normally distributed as well as for distributions that are not (see Appendix B for technical details about the various procedures). There are even procedures that can be used to approximate the AUC when only a single pair of hit and false alarm rates are obtained (McNicol, 1972).

Regardless of which procedure is deemed most appropriate, each procedure measures the same thing – the degree of overlap between underlying probability distributions. In addition, each procedure provides scores that range in value from 0.50 to 1.00 making them comparable. There are other possible indices that also measure distribution overlap (e.g., Guttman's DISCO), but none seem to have the theoretical and practical advantages that the AUC measure has (Swets, 1986).

2.6. Diagnostic decision-making

Building on the research by signal detection theorists, ROC analysis has also been used in diagnostic fields where two alternative, yes-no type decisions are frequently encountered. For example, radiologists use ROC analysis to assist with their determinations of whether a tumour has spread or not (Partin *et al.*, 1997). Meteorologists use it to assist with their predictions of whether serious

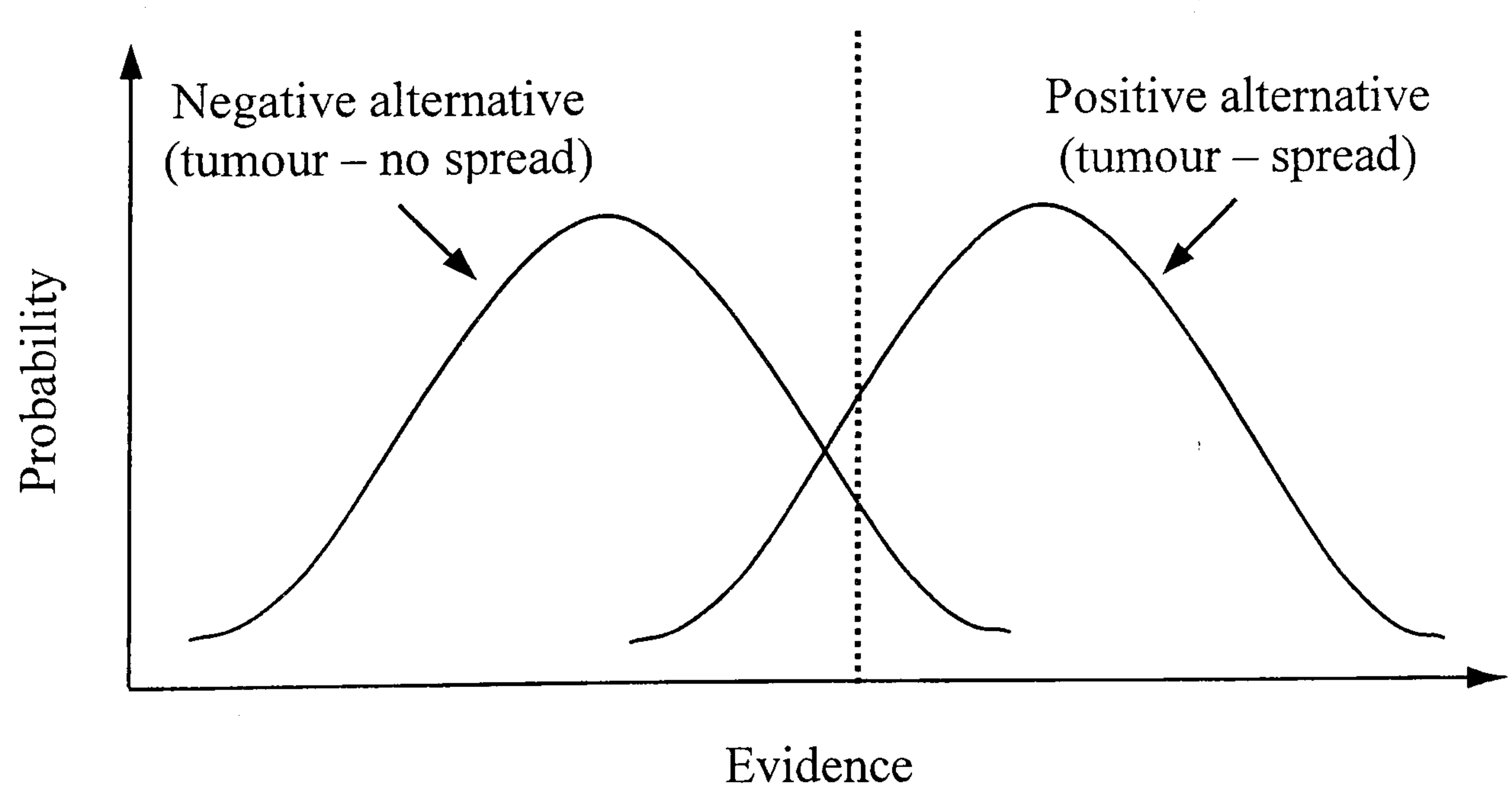
---

[1]To get some idea of what the AUC might mean in practice, it may be helpful to consider that it corresponds to the percentage of correct decisions made in a 2 alternative forced choice experiment (2AFC). A typical 2AFC experiment would consist of presenting a participant with 100 pairs of stimuli, where each pair consists of a randomly selected signal and a randomly selected noise. The task for the participant is to determine, on each of the 100 trials, which stimuli is which. The percentage of trials where the participant makes a correct decision corresponds to an AUC developed from the signal-noise distributions (Swets, 1988).

storms are going to strike (Carter & Polger, 1986). Psychologists use it to decide whether offenders are going to pose a significant threat to the community once they are released from prison (Steadman *et al.*, 2000). Engineers use it to help detect cracks in airplane wings (Swets, 1992). Some of the other diagnostic fields where ROC analysis has now been used include the forensic sciences, the military, education, medicine, economics and accounting.

The rationale behind using ROC analysis to assist with these decisions is that the two alternatives are analogous to the signal and noise faced by the radar detector discussed in the previous section. In line with this, the diagnostic alternatives are usually thought of as overlapping probability distributions. Consistent with the line of thinking presented thus far, one distribution represents a negative diagnostic alternative (e.g. a tumour that has not spread) and the other represents a positive diagnostic alternative (e.g. a tumour that has spread) (see Figure 2.5). In addition, the goal in these tasks is ultimately the same as it is in the signal detection task. That is, to detect a signal amongst a background of random interference or noise.

Figure 2.5. Hypothetical distributions representing two diagnoses



2.6.1. Increasing the accuracy of diagnostic decisions

ROC analysis has been used in two ways within diagnostic fields. The first use has been to increase diagnostic accuracy. That is, "...to enhance the odds that any given decision will be the correct one" (Swets, Dawes & Monahan, 2000b, p. 82). This first goal is typically accomplished by using various statistical techniques to identify cues that are best able to discriminate between the diagnostic alternatives of interest. In essence, the techniques allow one to identify features that are reliably associated with one alternative but not the other, thus identifying signal-noise distributions that overlap to a small degree. ROC analysis can assist with finding these features because, as already pointed out, the degree of overlap in these distributions is reflected in the height of the ROC curve.

As an example, consider the common radiological task presented above, where a radiologist must determine whether a tumour has spread to other parts of the body. In the case of prostate cancer, there are a variety of cues that can help with such a diagnosis (Partin *et al.*, 1997). The question for the radiologist quickly becomes, which of these cues will lead to the most accurate diagnosis? If one has a sample of patients where the diagnostic outcome is already known, and all of the potential cues have been measured in each patient, it is a relatively simple task to use ROC analysis in order to identify the cues that result in accurate predictions.

2.6.2. Increasing the utility of diagnostic decisions

The second way that ROC analysis has been used within the diagnostic fields is to increase the utility of diagnostic decisions. That is, to ensure "...that the number of true cases found does not come at the cost of an unreasonable number of false positive diagnoses" (Swets *et al.*, 2000b, p. 82). Utility can be enhanced by selecting an operating point along the ROC curve that provides the desired balance between the possible decision outcomes for the situation at hand. There are now a number of relatively well-known methods for selecting appropriate decision thresholds and comprehensive reviews of these methods do exist (e.g., Cantor, Sun, Tortolero-Luna, Richards-Kortum, Follen, 1999; Greiner, Pfeiffer & Smith, 2000; Swets, 1996). The four most common methods will be briefly reviewed here (see Table 2.3 for a summary and Appendix A for all relevant probability calculations).

Table 2.3. Methods for selecting an appropriate decision threshold

Method	Definition	Formula
Optimal	Identifies an appropriate decision threshold by combining the prior probabilities of the diagnostic alternatives with the costs and benefits associated with incorrect and correct diagnostic decisions.	$\frac{p(-)}{p(+)} \times \frac{B(\mathrm{CR}) + C(\mathrm{M})}{B(\mathrm{H}) + C(\mathrm{FA})}$
Ratio	Identifies an appropriate decision threshold by combining a ratio of the prior probabilities of the diagnostic alternatives with a ratio of the costs and benefits associated with incorrect and correct diagnostic decisions.	$p(-) : p(+) \times (B + C)(-) : (B + C)(+)$
Maximisation	Identifies an appropriate decision threshold by identifying the point where the maximum number of hits and the minimum number of false alarms are made.	$(p\mathrm{H} + p\mathrm{CR}) - 1$
Pre-selection	Identifies an appropriate decision threshold by ensuring that a pre-selected rate of false alarms (or hits) is not violated.	N/A

$p(-)$=prior probability of the negative diagnostic alternative occurring; $p(+)$=prior probability of the positive diagnostic alternative occurring; B(CR)=benefits associated with a correct rejection; C(M)=costs associated with a miss; B(H)=benefits associated with a hit; C(FA)=costs associated with a false alarm; $(B+C)(-)$=benefits and costs associated with a negative diagnostic alternative; $(B+C)(+)$=benefits and costs associated with a positive diagnostic alternative; pH= probability of a hit; pCR=probability of a correct rejection

(a) The optimal method

The most effective method for selecting an appropriate decision threshold is to consider: (1) the prevalence of the two alternatives in the target population (e.g., the probability of actually getting an unlinked crime pair and a linked crime pair), and (2) the costs and benefits associated with incorrect and correct decisions outcomes respectively (Peterson *et al.*, 1954). Using the optimal formula presented in Table 2.3 to combine this information one can identify the optimal operating point along a ROC curve, that is, a decision threshold that will lead to optimal discrimination performance.

(b) The ratio method

When it is not possible to estimate each of the values in the formula for the optimal method, other methods exist whereby ratios can be used instead. For example, it might be possible to estimate the prior probabilities of the diagnostic alternatives but it may be very difficult to assign specific costs and benefits to the various decision outcomes. This would especially be the case when peoples' lives or rights are at stake. However, as Swets and his colleagues (2000a) point out, we may still know that we would rather be right twice as much when the positive alternative exists as when the negative alternative exists. In this case we can insert this ratio (1:2) into the formula presented in Table 2.3 to determine which decision threshold to use.

(c) The maximisation method

There are also methods that exist when one wishes to ignore prior probabilities and the costs and benefits associated with each decision outcome. One of these methods results in the diagnostician making the maximum number of hits while making the minimum number of false alarms. This can be accomplished using a decision threshold falling along the ROC curve at a point closest to the upper left corner of the ROC graph. Using the maximisation formula presented in Table 2.3 it is possible to precisely determine what this operating point is.

(d) The pre-selection method

Alternatively, the diagnostician can pre-select a fixed rate of false alarms or hits that is deemed appropriate for the situation at hand and avoid making any diagnostic decision that violates this pre-selected rate. A police force may decide, for example, that the resources they have available for burglary investigations will prevent them from ever being able to exceed pFA=0.20. What is an appropriate rate of hits or false alarms would still depend on the relative costs and benefits associated with incorrect and correct decisions, but it would no longer be necessary to define these costs and benefits in a very precise way. No formula is required for this pre-selection option.

2.7. ROC analysis in the criminal context

The discussion so far has focused on where ROC analysis came from and why, as well as providing some insight into a few key features associated with the analytical procedure. The emphasis now will be on presenting ROC analysis as a potential framework for examining behavioural consistency and discrimination in serial burglary. In addition, some of the more detailed and technical aspects of the procedure, including how to actually construct an empirical ROC curve, will be presented.

2.7.1. Using ROC analysis to examine behavioural consistency

The logic behind using ROC analysis to examine behavioural consistency in serial burglary lies in the fact that the various behavioural processes underlying consistency and discrimination can be represented as overlapping probability distributions of across-crime similarity scores (see Chapter 1). It will be recalled that if offenders exhibit high levels of consistency, this will be reflected in underlying probability distributions that overlap very little. In contrast, if offenders exhibit low levels of consistency, this will be reflected in underlying probability distributions that overlap very much. If this is accepted as true, it follows that different levels of behavioural consistency should have characteristic ROC curves associated with them.

If intra-offender and inter-offender distributions overlap almost completely this will result in a ROC curve that is relatively low in its ROC graph. If the distributions overlap less this will result in a ROC curve that is relatively high in its ROC graph. It is important to note, however, that the height of a ROC curve does not depend in any way on the relative position of the distributions, that is, whether one distribution is to the right or the left of the other along the x-axis. The height of the ROC curve is purely a function of distribution overlap. Indeed, there are cases where intra-offender and inter-offender distributions would be expected to swap places in the present research because the relative position of each distribution depends totally on the nature of the evidence examined and the type of similarity measure used.

2.7.2. Using ROC analysis to examine behavioural discrimination

In addition to providing a measure of behavioural consistency, the height of the ROC curve can reflect overall levels of behavioural discrimination. There is an obvious relationship between consistency and discrimination, such that as the level of consistency increases within a sample so to does the level of discrimination that will ultimately be possible. In other words, if intra-offender and inter-offender distributions of similarity scores overlap very little, the degree of consistency would be high and high levels of discrimination would be possible. On the other hand, if intra-offender and inter-offender distributions of similarity scores overlap completely, the degree of consistency would be low and low levels of behavioural discrimination would be possible.

Beyond this general measure of behavioural discrimination, however, is the equally important fact that ROC analysis allows discrimination accuracy to be assessed at each and every decision threshold. This can be done simply by examining the probability of each decision outcome resulting at different operating points falling along a ROC curve. In addition, using one of the strategies presented in Table 2.3, or others that exist in the literature, decision thresholds can be identified that result in a desired balance between the four possible decision outcomes.

2.7.3. Constructing an empirical ROC curve in the criminal context

In order to create empirical ROC curves, like the one illustrated in Figure 2.4, values of pH and pFA need to be calculated across various decision thresholds. In a typical signal detection experiment, this is a relatively simple task. For example, it could be accomplished by instructing the observer to change their decision threshold from trial to trial, from strict to lax, for a given signal-noise distribution (McNicol, 1972). For each decision threshold, the frequency of hits, misses, correct rejections and false alarms can be calculated and converted into the conditional probabilities of interest, resulting in the same number of ROC points as there are different thresholds. These points could then be plotted on a ROC graph to give a single ROC curve.

Alternatively, the observer can be instructed to state their decisions in terms of how probable they think it is that a signal was present, say on a scale from 0 to 100. In effect, what this means is that the observer is setting multiple decision thresholds simultaneously (McNicol, 1972). The researcher could then impose a number of cut-off points along this continuous output, calculate pH and pFA at each one of these points, and plot the data on a ROC graph to give a single ROC curve.

When an actual observer or decision-maker is not involved in the process, as will be the case in the present research, an empirical ROC curve can be just as easily created from the continuous output of an analytical procedure (Swets *et al.*, 2000a). In this case, statistical or computational techniques such as regression analysis (Hosmer & Lemeshow, 1989), decision tree analysis (Hawkins, 1982), discriminant function analysis (Lachenbruch, 1975), or artificial neural networks (Hertz, Krogh & Palmer, 1991) can provide probabilities that a given alternative has occurred on the basis of some sort of evidence.

In the present case, the evidence provided to these procedures might be across-crime similarity scores resulting from a particular set of crime scene behaviours. The alternatives of interest will be whether the similarity scores originated from intra-offender or inter-offender observations. As when using probability ratings provided by actual observers, the researcher can impose multiple thresholds at various points along the continuous output. In this case, the threshold may range from very strict (e.g., $p \geq 0.90$) to very lenient (e.g., $p \geq 0.10$) using intervals of 0.10. Values of pH and pFA can then be calculated for outputs that exceed each successive threshold, resulting in numerous ROC points that can be plotted to give a single ROC curve.

As in the signal detection task, to get another ROC curve in the criminal context, the evidence provided to the analytical procedure would simply need to be altered in some way. For example, across-crime similarity scores based on a different set of crime scene behaviours could be treated as the evidence and the process could be repeated. In this way, the height of the resulting ROC curves, as measured by their AUCs, could be used as an indication of the conditions that lead to the highest levels of consistency and discrimination in serial burglary.

2.7.4. Advantages of using measures derived from the ROC curve

In order to evaluate the extent to which offenders exhibit behavioural consistency and discrimination in serial burglary, it is necessary to have an appropriate measure. As indicated at the beginning of this chapter, this measure should represent the overall degree of overlap between intra-offender and inter-offender similarity scores and the probability of making various decisions using different levels of across-crime similarity. The ROC curve, and its associated AUC, is able to represent these two separate criteria. Indeed, there are several advantages associated with using ROC-related measures as a way of quantifying consistency and discrimination.

(a) The AUC provides a single measure

One of the most obvious advantages of using the AUC is simply that it summarises a vast amount of information in a single quantitative score. As in the signal detection context, this is the case because the AUC is independent of any specific decision threshold. This provides, amongst other things, a way of objectively identifying those aspects of offence behaviour that are most consistently exhibited by a sample of offenders. It also provides a way of identifying those aspects of offence behaviour that should be used for discrimination purposes.

(b) The AUC provides a flexible measure

A second advantage of the AUC is its flexibility, in that it can be easily used to summarise levels of consistency and overall discrimination regardless of what aspect of offending behaviour is observed or how across-crime similarity is measured. For example, crime site locations and property stolen are different behavioural domains and different procedures are used to indicate if they are expressed in a similar fashion across crimes. Nevertheless, the observed consistency levels across these two domains can be compared using the AUC because the measure is not domain or procedure dependent.

(c) The AUC provides a general measure

A third advantage of the AUC is that it can be compared across samples that differ in various ways (Swets *et al.*, 2000a). This is because the AUC is based solely on the proportions of hits and false alarms rather than the relative frequencies or base rates of the possible alternatives. As a result, the AUC is not specific to any particular sample and can be used to compare consistency or discrimination levels across police jurisdictions. This would be the case even if those jurisdictions differed substantially in the rates of intra-offender or inter-offender crime they experience.

(d) The ROC provides a threshold-specific measure

A fourth advantage of using ROC-related measures in this context is that the ROC curve indicates the level of discrimination possible at each and every decision threshold. This is in line with the fact that discrimination accuracy is a function, not only of distribution overlap, but also of where the decision threshold is placed. Until now, the placement of decision thresholds in investigative tasks such as comparative case analysis has been decided in a purely arbitrary fashion. There has been no appreciation of what an appropriate decision threshold may look like or how pH and pFA will vary as the threshold is changed (e.g., Canter & Heritage, 1991; Grubin *et al.*, 2001).

2.8. A note regarding the limitations of the ROC framework

Throughout this chapter, there have only ever been two alternatives to decide between in each of the tasks discussed. In the discussion of psychophysics, the decision was between stimulus A and stimulus B. In statistical hypothesis testing, the decision was between the null hypothesis and the alternative hypothesis. In signal detection theory, the decision was between background noise and a signal. In diagnostic decision-making, the decision was between a tumour that has spread and one that has not. Lastly, in the discussion of consistency and discrimination in the criminal context, the decision was between an inter-offender observation of behaviour and an intra-offender observation.

While seemingly a minor point, this is a crucial one for a variety of reasons. The primary reason is because ROC analysis is only capable of dealing with yes-no type decisions. At present, there exists no generalised form of ROC analysis that can deal with decisions involving multiple alternatives, primarily due to the complexities underlying such tasks (Metz, Starr & Lusted, 1977). This fact is important, in turn, because it determines the sorts of statistical techniques that can be used to study yes-no type tasks. Despite all of this, yes-no type decisions often exist and are extremely important. Even tasks involving multiple decision alternatives can usually be broken down into a sequence of two alternative decisions (Swets & Pickett, 1982).

CHAPTER 3
METHODOLOGICAL AND ANALYTICAL ISSUES

3.1. Introduction

As a result of this thesis being organised such that similar data and analytical procedures are used in each successive chapter, certain methodological and analytical issues will run throughout the research. It therefore makes sense to discuss these issues up front, before presenting the results of any analysis.

3.2. Methodological issues

A number of methodological issues will be discussed. These include issues about the type of data used in the present research, where the data was collected, how the data was collected, and some possible benefits and limitations associated with the data.

3.2.1. The participating police forces

A number of police forces throughout the UK were the primary suppliers of data for the present research. One of the forces covered a large, mostly rural area (Dorset) while the others were predominantly urban police forces (London, Greater Manchester (Oldham division) and Merseyside). In each case, solved serial burglaries were focused on. This was done for two reasons. First, burglary data is one of the most widely available sources of police data and the most readily accessible. Second, the police have a particular problem investigating burglaries, primarily due to their very high frequency of occurrence (Clarke & Hough, 1980; Heal & Morris, 1981).

Where possible, data was collected on both residential and commercial burglaries[2]. However, in two of the police forces, only residential burglaries could be provided (London and Dorset). In addition, an attempt was made to collect data on as wide a range of burglary behaviours as possible. While no comprehensive model can be located in the published literature that describes the

[2]Residential burglaries are defined as any burglary where an offender targets a domestic dwelling. Commercial burglaries are defined as any burglary where the offender targets a commercial property.

various components of burglary behaviour, breaking down burglaries into spatial behaviour, target selection choices, entry behaviours, property stolen and internal behaviours seems to exhaust the range of possible actions. In the majority of cases, each of these aspects of burglary behaviour was collected. However, in the case of one police force, only spatial behaviour could be provided (London). Table 3.1 on the next page, provides a brief summary of the data collected from each police force. A more detailed list of the specific offence behaviours collected from each force can be found in the variable lists making up Appendices E through I.

3.2.2. The data collection process

The data collection procedure adopted for this research was the same regardless of the police force. The procedure consisted of five general stages: (1) getting permission to collect the data, (2) becoming familiar with the data available, (3) becoming familiar with data recording and storage procedures, (4) extracting the relevant data, and (5) ensuring anonymity within the data.

(a) Getting permission to collect the data

Each police force was contacted with a request for behavioural information relating to relatively recent, solved residential and commercial burglary offences. Authorisation was provided in each case, though restrictions prevented the collection of detailed offender information, including in some cases the home locations of offenders. Data collection took place at the headquarters of each police force. The exception to this was the spatial data provided by London. This data was collected by crime analysts working within that police force at the time the offences were committed and was provided specifically to examine the spatial behaviour of residential burglars committing crimes in London.

Table 3.1 A brief summary of the burglary data

Sample	Size	Crime	Time	Behaviours
London	69 offenders (816 offences)	Residential	1999-2000	Spatial behaviour
Dorset (Appendix E)	28 offenders (233 offences)	Residential	1997-1998	Spatial behaviour Entry behaviour Targeting behaviour Property stolen Internal behaviour
Oldham 1 (Appendix F)	36 offenders (150 offences)	Residential	1999	Spatial behaviour Entry behaviour Targeting behaviour Property stolen
Oldham 2 (Appendix G)	43 offenders (135 offences)	Commercial	1999	Spatial behaviour Entry behaviour Targeting behaviour Property stolen
Merseyside 1 (Appendix H)	51 offenders (660 offences)	Residential	1995-1999	Spatial behaviour Entry behaviour Targeting behaviour Property stolen Internal behaviour
Merseyside 2 (Appendix I)	57 offenders (634 offences)	Commercial	1994-1999	Spatial behaviour Entry behaviour Targeting behaviour Property stolen Internal behaviour

(b) Becoming familiar with the data available

Upon initially arriving at each police force, it was first necessary to ensure there was sufficient behavioural material available for use. This consisted of reviewing the contents of crime reports to ensure the desired information was routinely collected. The crime analysts responsible for inputting burglary data at each police force were also regularly consulted to get their opinion on whether the available data could be used to examine the issues focused on in the present research. A sufficient amount of information was found in each of the forces. Additionally, while each police force rarely recorded the exact same information, there was found to be a high degree of general overlap in the sort of material collected. Even across residential and commercial burglary, broad similarities existed (across crime type and across police forces) in the sort of information recorded.

(c) Becoming familiar with data recording/storage procedures

Having ensured that sufficient data was available within each police force, data collection began. At this stage, it was necessary to become familiar with the various data recording/storage procedures used in each force. In all three forces, this involved some basic training provided by a crime analyst familiar with the procedures. Each of the police forces was found to carry out similar procedures, which began with an investigating officer filling out a detailed crime report while present at the crime scene. These crime reports typically consisted of a pre-defined checklist of variables relating to various offence characteristics that the officer would record as being present or absent. This was typically followed by space for the officer to provide a written summary of the offence. Little specialised training seemed to be provided to these officers with respect to how they should go about recording information.

Once the investigating officer had recorded the offence information, the report was passed to a crime analyst who input the data into a computerised database for storage and future analysis. Each police force used a different storage system for this purpose, which had implications for how data could be retrieved in the Dorset research. Dorset burglary data is entered into the Dorset Crime Database (DCD) where it is possible to retrieve information about each burglary either in free-text format or as a list of variables dichotomously coded as being present or absent. In Oldham, where burglary data is entered into the Crime Pattern Analysis (CPA) package, it is also possible to retrieve data from the database in either free-text format or as a list of dichotomously coded variables. In Merseyside, on the other hand, where burglary data is entered into the Integrated

Criminal Justice System (ICJS), it is only possible to retrieve data in free-text format.

(d) Extracting the relevant data

The next stage of the data collection procedure was to extract the relevant behavioural material. When the data was retrievable as a list of dichotomously coded variables this greatly facilitated the process. In these cases, all the data could be directly exported from the database into a more conventional, user-friendly software application. Any additional information beyond what was included in the list of dichotomous variables could then be added at this time, by content analysing the free-text portion of crime reports. A note was kept about what data was coded from free-text in case problems existed with its reliability.

With the Merseyside data, the procedure was more time consuming. In this case, because the data was only retrievable in free-text format, hard copies of the text had to be printed off and content analysed. The variable lists used for this purpose were based on the burglary behaviours collected from the other police forces (see Appendix H and I). This process was facilitated by utilising standard software designed to automate the data collection process (e.g., SPSS Data Builder). The end result for each data set was a data matrix consisting of offences as rows and dichotomously coded crime scene behaviours as columns, in addition to two columns containing geo-coded x and y coordinates indicating the position of each crime site location.

(e) Ensuring anonymity within the data

Once all of the data was downloaded and in a form suitable for statistical analysis, the final stage of the collection procedure was to ensure all data was made sufficiently anonymous. To achieve this, all identifying features were removed from the data, including the names, addresses and phone numbers of offenders and victims. Data was replaced by a simple coding system whereby an identification number was assigned to each offender and each one of their crimes (e.g., the first code, 1-1, corresponds to offender 1 - burglary 1). The data was then approved by the individual within each police force overseeing data collection to ensure that the confidentiality requirements of the force were met.

3.2.3. Potential benefits and limitations of using police data

Examinations of behavioural consistency and discrimination in the non-criminal context are almost always based on direct observations of people's behaviour, or on reports from those people who observe it first hand (e.g., Bem & Allen, 1974; Funder & Colvin, 1991; Mischel & Peake, 1982; Shoda, Mischel & Wright, 1994). In contrast, in the criminal context the same issues must be examined through the use of second hand data collected by the police, since offence behaviour is rarely ever directly observable.

Police data can come in a wide variety of forms (e.g., police reports, interviews with offenders, victim statements, eyewitness accounts, crime scene photographs, etc.) and it tells the researcher much about the way in which crimes are committed. However, there are potential problems inherent in the use of such material and therefore questions will inevitably arise over its suitability for research. Neither the benefits nor the limitations of using police information as the sole source of data in psychological research should be overlooked, so a brief review of both is provided below.

3.2.4. Potential benefits of using police data

There are a number of potential benefits associated with using police data for psychological research. Two primary benefits include the fact that: (a) police data is not influenced by the research agenda to a great extent, and (b) police data is ecologically valid and practically relevant.

(a) Police data is largely unaffected by the research agenda

The participants in studies of non-criminal consistency usually know they are being studied. In fact, many of the studies take place in a laboratory setting (e.g., Funder & Colvin, 1991). While the artificial nature of the laboratory situation does not necessarily mean the results from such studies are meaningless, one must question the extent to which being observed under such conditions effects participant behaviour (Lee, 2000). In other words, although participants may exhibit reasonable levels of consistency, there is no way to know how much of this consistency can be attributed to the person being observed and how much can be attributed to various sorts of experiment and/or experimenter effects.

In the criminal context, the story is very different. The researcher in this case usually has nothing, or certainly very little, to do with the data collection procedure. While a number of problems may result as a consequence of this, one of the significant benefits is the lack of interference from outside sources. One can be fairly confident that the behaviour an offender exhibits during the commission of his crimes is primarily a result of factors occurring naturally in that context. Therefore, any behavioural consistency that exists across an offender's crimes, and any discrimination that is possible as a result, is likely to be a true reflection of how consistent that offender is. The real challenge with measuring consistency and discrimination, then, is to establish a suitable method for accurately recording what offenders do. If this can be accomplished, police data can offer a useful source of information for studying behavioural consistency and discrimination in the criminal context.

(b) Police data is ecologically valid and practically relevant

By relying on police data, any findings that emerge from the present research can claim some important ecological validity and consequent practical relevance. Indeed, many of the findings from this research will have very immediate and direct practical significance because they are based, quite literally, on operational databases. That is, a number of the databases used in the present research are identical to those the police have to work with in their investigations of serial burglary (e.g., the Oldham database). In this sense, it may be important that the data contain some of the potential problems outlined below, because until the police adopt more effective data collection procedures these problems will exist and have to be dealt with by law enforcement officers.

3.2.5. Potential limitations of using police data

In addition to the potential benefits associated with the use of police data, there are also possible limitations. This is the case especially when using data from solved crimes to examine behavioural consistency and discrimination in the criminal context. These limitations include the fact that: (a) such data may not provide material that is representative of crimes or criminals in general, (b) such data may not provide an accurate account of what actually happens in a crime, and (c) such data may not indicate with certainty whether an offender is responsible for the crime in question.

(a) Police data may be unrepresentative

It is now a well-known fact that crimes often go unreported, that not all crimes reported to the police are recorded as crimes, and that even a smaller percentage of these recorded crimes are ever solved (Coleman & Moynihan, 1996). In order to examine issues such as behavioural consistency and discrimination in the criminal context, however, research needs to be based on serial crimes where the offender is known. Thus, solved crimes end up forming the bulk of most research databases, creating a situation where research is based on incomplete and possibly unrepresentative samples of crime.

The extent to which the results generated from such research can be generalised to a wider population of crimes and criminals must be questioned. It must be accepted, for example, if high levels of behavioural consistency are in fact found that this might be a feature that is particular to solved crimes. Indeed, high levels of consistency might be one of the significant reasons why these crimes become solved in the first place. For example, the fact that two crimes by the same offender are committed in roughly the same area at roughly the same time may significantly increase the chance of those crimes being linked and solved. However, a third crime committed by the same offender may remain unsolved if it is more geographically and temporally dispersed.

A by-product of this problem is that there may be gaps within any offence series, making it unlikely that all crimes committed by each offender will be included in any research (Grubin *et al.*, 2001). Large time gaps between crimes may have a serious impact on observed levels of behavioural consistency, and consequently on the level of discrimination that is possible. For example, it is possible that an offender may display high levels of consistency across crimes committed relatively close together in time, but considering just two of his offences committed two years apart may substantially decrease his level of consistency. Not only would natural processes like learning and maturation have more impact on an offender's behaviour over long periods of time, important environmental factors might also substantially change during this time. A commercial burglar's spatial behaviour may change over time, for example, as a result of new commercial sites being built or road systems being developed or removed.

(b) Police data may be inaccurate

Beyond the issue of whether police data from solved crimes is representative of crimes in general, there are limitations relating to the accuracy of such data. At times, the police are known to record crime scene information in an inaccurate fashion (Farrington & Lambert, 1992). Two of the most obvious reasons for these inaccuracies include the variations that exist in collection procedures and factors of distortion that exist amongst people involved in the collection process (Alison, Snook & Stein, 2001).

Problems can result in many police forces because investigating officers are simply instructed to provide a narrative account of how crimes are committed, while being offered little in the way of specific guidelines regarding what should be collected or how (Alison *et al.*, 2001). In such situations, it is likely that variations will exist across the accounts provided by different investigating officers, or even within the accounts provided by the same officer over time. These variations may have nothing to do with how crimes have been committed. For example, something as seemingly simple as recording whether a burgled property is well-maintained at the time of the offence can create significant problems, if what is meant by well-maintained is not made totally clear. All that has to happen for a consistent offender (with respect to his target selection choices) to suddenly become less consistent is for investigators to have different interpretations of what well-maintained means[3].

Alison *et al.* (2001) also point out that many inaccuracies can creep into police data due to the fact that every person involved in the collection process (victims, witnesses, police, prosecutors, etc.) typically has a different agenda. Essentially, each person giving an account of the crime will have his or her own reason for

[3]This particular example is used because it was presented as a significant problem at a recent National Burglary Analysis Conference hosted by the Metropolitan Police Service in 1999. Other, similar examples were also presented as problematic.

providing that account, and each account will vary depending on the motive for giving it and as a function of whom the account is being given to. When reporting a burglary to the police, for example, a victim may leave out or add in information because their primary motive for giving the account may be to make a successful insurance claim. The police, on the other hand, may leave out or add in information because their primary motive for giving the account is to provide a convincing case to the prosecution service.

Variations in police data can reflect actual variation that exists across crimes. However, when there is no strict protocol in place specifying how data should be collected it is difficult to know whether this is the case. One simple way of ensuring this problem is reduced is to develop more systematic collection procedures that officers can use, perhaps in the form of carefully constructed checklists (Lee, 2000). While this may not help with certain forms of distortion, it should reduce the variations caused by inappropriate collection procedures. As a way of trying to maximize the quality of data used in the present research, only police forces using this structured procedure were targeted for data.

(c) Police data may not reflect criminal responsibility

A third limitation with police data relates to how confident we can be that an offender arrested for a particular crime is in fact responsible for the crime in question. This potential problem is worthy of discussion for at least two reasons. The first reason is that the present research totally depends on knowing which offenders in the sample are responsible for what crimes. If this cannot be determined with a fairly high degree of confidence, behavioural consistency is unlikely to be found and only low levels of discrimination accuracy will ever be achieved.

The second reason it is important to mention this problem relates to a potentially biasing factor existing within police forces where there is pressure to increase clear-up rates. One common method for increasing clear-up rates has been to get offenders, in exchange for more lenient treatment by the police and the courts, to admit to other crimes they have committed. These crimes are often referred to as crimes that are 'taken into consideration', or TIC's. While TIC's need not represent a serious problem, the extent to which the practice has been abused has recently been disclosed (The Guardian, March 18, 1999). Essentially, some offenders are being assigned crimes they did not and often could not commit in an attempt to meet overly ambitious target rates.

The problem for the present research is that the basis for inappropriate TIC's may depend in some way on how consistent the crimes were in the first place. This would result in an artificial increase in the level of consistency observed. For example, because two crimes have been committed close together in space they may be viewed as crimes worthy of a TIC. Unfortunately, it is difficult to eliminate this problem totally. The only way to do this would be to start using DNA evidence as the sole basis for establishing guilt. While DNA evidence has been used for this purpose in a recent study (Wiles & Costello, 2000), it was not feasible to make this a criterion for accepting data in the present research.

3.3. Analytical issues

As mentioned at the beginning of this chapter, the same set of analytical procedures will be used throughout the upcoming chapters to examine each of the burglary samples. It will, therefore, be beneficial to provide an initial overview of these procedures and the various analytical issues that may arise. The process when analysing each sample of burglaries involves the same three general stages. First, a variety of similarity scores are calculated between each and every crime. Second, the similarity scores are used to create logistic regression models. Third, the probabilities generated from these models are subjected to ROC analysis in order to obtain separate measures of consistency and discrimination.

3.3.1. Calculating spatial and behavioural similarity scores

The dichotomous criterion variable in the present research is always related to whether the same offender or different offenders committed a pair of crimes. The predictor variables, on the other hand, are all continuous. One of these predictor variables relates to inter-crime distances, that is, the distance in kilometres between each and every crime. The other predictor variables relate to across-crime similarity scores that pertain to 4 different behavioural domains, including entry behaviours, target selection choices, property stolen and internal behaviours.

Each of the 5 possible predictor variables is based on the premise that a higher degree of behavioural similarity and distinctiveness will be exhibited across crimes committed by the same offender. Thus, it is expected that crimes committed by the same offender will be characterised by shorter inter-crime distances and higher across-crime similarity scores for entry behaviours, target selection choices, property stolen, and internal behaviours.

(a) *S-LINK and B-LINK*

Due to the large number of crime pairs that result from large samples of offences, two computer programs were developed to automate the calculation of spatial and behavioural similarity scores[4]. The program used to calculate spatial similarity scores will henceforth be referred to as *S-LINK* (Appendix C contains instructions for using this program). The program used to calculate behavioural similarity scores will henceforth be referred to as *B-LINK* (Appendix D contains instructions for using this program).

(b) Calculating spatial similarity scores

Spatial similarity scores are equivalent to the Euclidean distance between each and every crime. Two crimes are said to be spatially similar when the distance between those crimes is relatively small. Two crimes are said to be spatially dissimilar when the distance between those crimes is relatively large. In the present research, spatial similarity scores are calculated by inputting the geo-coded x and y coordinates indicating the position of each crime site location into *S-LINK*.

These geo-coded coordinates typically consist of 7 digits corresponding to the exact geographic location of each crime site to the nearest metre. Using Pythagorean theorem as the basis for all calculations, *S-LINK* provides as output

[4]The number of crime pairs resulting from a given sample of crimes can be calculated using: $C_r^N = \frac{N!}{r!(N-r)!}$, where C_r^N refers to the number of combinations of N crimes taken r at a time.

the distance in kilometres between every pair of crimes. An example output from *S-LINK* would look something like Figure 3.1. The first column represents the crimes making up each pair, the second column provides an indication of whether the crime pair is linked or unlinked, and the third column represents the inter-crime distance for each crime pair. Thus, the distance between the first two crimes committed by offender 1, a linked crime pair, is 0.10 km.

Figure 3.1. An example output from *S-LINK*

Pair	Linked	Distance
1-1, 1-2	Y	0.10 km
1-1, 1-3	Y	1.78 km
1-2, 1-3	Y	1.82 km
1-1, 2-1	N	8.60 km
1-1, 2-2	N	9.26 km

Note: 1-1: offender 1 - crime 1; 1-2: offender 1 - crime 2; etc.

(c) Calculating behavioural similarity scores

Jaccard's coefficient is used as the behavioural similarity measure for all other behavioural domains - entry behaviours, target selection choices, property stolen and internal behaviours. 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. Jaccard's coefficient ranges from 0 to 1. Two crimes are said to be behaviourally similar when the behaviors exhibited within those crimes are the same (i.e., coefficient values approaching 1). Two crimes are said to be behaviourally dissimilar when the behaviours exhibited within those crimes are different (i.e., coefficient values approaching 0).

As an example of how to calculate Jaccard's coefficient, consider two burglaries that have been dichotomously coded across 17 entry behaviours. A value of 0 indicates a behaviour that was absent and a value of 1 indicates a behaviour that was present. The pattern of entry behaviours in crime 1 is 0000000000001111 and in crime 2 it is 1100000001111111. If a equals the number of behaviours present in both crimes (i.e., 1/1), b and c equal the number of behaviours present in one crime but not the other (i.e., 0/1 or 1/0), and d equals the number of behaviours absent from both crimes (i.e., 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.

Considering the unverifiable nature of burglary data, and the distinct possibility that variables are not recorded as being present when they are in fact present, it may be useful to ignore joint non-occurrences when assessing across-crime similarity. Beyond this, the use of Jaccard's coefficient for this purpose is in line with previous examinations of behavioural discrimination in the criminal context (e.g., Canter & Heritage, 1991). Its use is also consistent with numerous other studies that have utilised police data in an attempt to identify patterns in offending behaviour (Bennell, Alison, Stein, Alison & Canter, 2001; Canter & Heritage, 1990; Canter & Fritzon, 1998; Canter, Hughes & Kirby, 1998; Salfati & Canter, 1999). However, it should be pointed out that Jaccard's coefficient is a relatively coarse-grained coefficient and therefore it may be useful in future research to develop a more refined similarity measure.

In order to calculate behavioural similarity scores, *B-LINK* was used. *B-LINK* takes as input a series of dichotomously coded variables pertaining to each behavioural domain. These indicate the presence or absence of the specific behavioural features making up these domains. For example, variables relating 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*). *B-LINK* then provides as output a similarity score between every pair of crimes for the entry domain, before providing output for the other domains. See Appendix E through I for variable lists corresponding to each burglary sample.

An example output from *B-LINK* would look something like Figure 3.2. The first column represents the crimes making up each pair, the second column provides an indication of whether the crime pair is linked or unlinked, and the remaining columns represent the behavioural similarity scores assigned to each pair for each domain. Thus, the level of similarity between the first two crimes committed by offender 1 is 0.78 for entry behaviours, 0.56 for target selection choices, 0.67 for property stolen, and 0.05 for internal behaviours.

Figure 3.2. An example output from *B-LINK*

Pair	Linked	Entry	Target	Property	Internal
1-1, 1-2	Y	0.78	0.56	0.67	0.05
1-1, 1-3	Y	0.62	0.32	0.45	0.15
1-2, 1-3	Y	0.59	0.49	0.49	0.23
1-1, 2-1	N	0.08	0.13	0.13	0.25
1-1, 2-2	N	0.32	0.32	0.18	0.18

Note: 1-1: offender 1 - crime 1; 1-2: offender 1 - crime 2; etc.

3.3.2. Developing logistic regression models

Once calculated, these spatial and behavioural similarity scores are used as the basis for developing logistic regression models. These models take the place of a real decision-maker in the present research and provide predicted probabilities that crimes are linked, which in essence are equivalent to discrimination decisions. It should be pointed out that a number of statistical procedures could have been used for this purpose. Examples of alternative procedures include discriminant function analysis, artificial neural networks, and decision tree analysis. Logistic regression analysis was chosen for four reasons.

First, logistic regression analysis can be used in situations where predictions need to be made about the values of a dichotomous criterion variable. Second, fewer assumptions need to be made when using logistic regression analysis compared to alternative techniques. For example, in contrast to logistic regression analysis, discriminant function analysis requires that predictor variables be normally distributed and that there is a linear relationship between the predictor variables and the criterion variable (Tabachnik & Fiddell, 1996). Third, unlike techniques such as artificial neural networks, logistic regression analysis is readily available in most statistical packages, easy to use, and relatively well understood. Fourth, logistic regression analysis is generally accepted as an analytical technique in a number of other diagnostic fields, such as clinical psychology (Steadman *et al.*, 2000), education (Swets *et al.*, 2000a), meteorology (Carter & Polger, 1986), and radiology (Getty *et al.*, 1997), where it is also commonly used in conjunction with ROC analysis.

(a) Maximum likelihood estimation

When carrying out simple linear or multiple regression analysis, regression coefficients are calculated through a process referred to as ordinary least squares estimation. Using this procedure, an attempt is made to minimise the sum of squared residuals. In logistic regression analysis, however, the same procedure cannot be used. Instead, the regression coefficients are estimated using a procedure known as maximum likelihood estimation (MLE) (Eliason, 1993). The MLE procedure is an iterative one. Parameter estimates are initially set and calculations are made to determine how well the parameter estimates fit the data. In the second iteration, the estimated parameters are changed slightly. If these changes result in a significantly better fit with the data the new estimates will be adopted, if not the previous estimates will be maintained. This procedure ends once the estimates cannot be improved.

(b) Definitions of the terms used in logistic regression analysis

In logistic regression analysis, log odds will often be referred to. In the present research, the log odds of a crime pair being linked can be calculated for any given sample of burglaries. For each sample, the log odds are expressed as a linear combination of variables. In the present context, these variables correspond to the various across-crime similarity scores. For a particular sample of burglaries, the combination of variables is given by an equation known as the logistic regression equation. The logistic regression equation is thus:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

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

Log odds can often be difficult to interpret, so they are often transformed into odds. When two crimes in a burglary sample are characterised by a particular across-crime similarity score, the odds of the crimes being linked is simply the probability that the two crimes are linked divided by the probability that the two crimes are not linked. To carry out this transformation, the log odds calculated for that burglary sample are simply exponentiated, as in:

$$\text{odds (linked)} = e^{\text{log odds}} = e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n}$$

If the odds were equal to 1 in this case this would suggest that a crime pair, given a particular similarity score, is just as likely to be linked as it is to be unlinked. In contrast, values of odds below 1 would suggest that a crime pair is more likely to be unlinked and values of odds above 1 would suggest that a crime pair is more likely to be linked.

Perhaps making everything even easier to understand, the odds can be converted into an estimated probability that two crimes are linked. As always, this probability can range from 0 to 1. If the data has been coded such that a value of 1 indicates a linked crime pair, a probability of 0 would indicate that a crime pair is not linked. In contrast, a probability of 1 would indicate that a crime pair is linked. Again, these probabilities are very simple to calculate by simply dividing the odds by 1 plus the odds:

$$p(\text{linked}) = \frac{\text{odds}}{1+\text{odds}} = \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n}}{1+e^{\alpha+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n}}$$

(c) Calculating the odds of two crimes being linked

As an example of how to calculate the odds of two crimes being linked, consider a logistic regression equation constructed for a particular sample of solved serial burglaries that consists of numerous linked and unlinked crime pairs. The logistic regression equation takes the form:

$$\log \text{odds} = -1.31 - 0.89X_1$$

where X_1 represents the inter-crime distances calculated between each and every crime in the sample. If two crimes in this sample are 2.0 km apart, the odds that those crimes are linked can be calculated using the above formula in the following way:

$$\log \text{odds} = -1.31 - 0.89(2.0) = -3.09$$

$$\text{odds} = e^{-3.09} = 0.05$$

Thus, the odds of these two crimes being linked, given that they are 2.0 km apart, is very low indeed.

Alternatively, consider two crimes in this sample that are only 1.0 km apart. In this case, the odds of the crimes being linked equals:

$$\log \text{odds} = -1.31 - 0.89(1.0) = -2.20$$

$$\text{odds} = e^{-2.20} = 0.11$$

Thus, in this hypothetical burglary sample, as the distance between two crimes decreases from 2.0 km to 1.0 km, the odds that the crimes are linked increases from 0.05 to 0.11. Findings such as these would confirm that crimes committed by the same offender within the hypothetical burglary sample are in fact characterised by shorter inter-crime distances.

The effect of changes in one or more of the predictor variables can also be examined by showing how a change of 'x' units effect the odds that crimes are linked (Hosmer & Lemeshow, 1989). For example, using the same logistic regression equation presented above, if the distance between two crimes in that sample increases by 1.0 km, the odds of the crimes being linked would be multiplied by:

$$e^{(x \times \beta)} = e^{(1.0 \times -0.89)} = 0.41$$

Multiplying odds by 0.41 would reduce them, which is consistent with the above example where the odds of two crimes being linked decreases from 0.11 to 0.05 as the inter-crime distance increases from 1.0 km to 2.0 km (e.g., $0.11 \times 0.41 = 0.05$).

However, changes in the predictor variables do not necessarily have to equal 1. Indeed, changes of 1 unit may not make sense with some of the similarity scores used in the present research. In the case of a similarity score based on Jaccard's coefficient, for example, it may make more sense to examine the effect of changes on the order of 0.10, considering that Jaccard's coefficient only ranges from 0 to 1. Thus, values of 'x' have to be considered in relation to the specific similarity score being examined.

(d) Calculating the probability of two crimes being linked

As already indicated, an alternative way of examining the effect of changes in one or more of the predictor variables is to examine what happens to the estimated probability of two crimes being linked when levels of across-crime similarity change. For example, given the two previous scenarios, where the first pair of crimes was 2.0 km apart and the second pair was 1.0 km apart, estimated probabilities can be calculated using the odds calculated above.

For the case where crimes are 2.0 km apart:

$$p(\text{linked}) = \frac{0.05}{1+0.05} = 0.05$$

and for the case where the crime are 1.0 km apart:

$$p(\text{linked}) = \frac{0.41}{1+0.41} = 0.29$$

Thus, as is the case with the odds of being linked, this increase in probability as inter-crime distances decrease would indicate that crimes committed by the same offender within this sample are characterised by shorter inter-crime distances.

(e) The relationship between log odds, odds and probabilities

When working with log odds, odds and probabilities 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. Which values are used, then, is simply a matter of preference. In line with this, odds or probabilities are the values typically used since they seem to be the values that are most easily understood.

The second point is that all these values are effected by how often the alternatives of interest occur. Since linked crime pairs will typically be very rare compared to unlinked crime pairs it should come as no surprise when linked crime pairs are associated with relatively low probabilities (or odds). This is not unusual, nor is it specific to the current task. The same thing would be found whenever a positive diagnostic alternative is rare compared to a negative diagnostic alternative (e.g., when diagnosing a rare form of cancer). What is important in this case are not the values of these probabilities per se, but rather how the probability of two crimes being linked compares to the probability of two crimes being unlinked.

For example, it may be the case in a particular burglary sample that most offenders are rarely known to have committed more than two burglaries. As a result, when crime pairs are constructed, unlinked crime pairs will out number linked crime pairs by a large margin. Under these circumstances, we should not expect any crime pair to have a high probability of being linked. Indeed, we may find that no linked crime pair has a probability exceeding 0.10. Looking at this probability in isolation would probably lead to an erroneous decision that the crime pair is unlinked. However, if the probabilities associated with unlinked crimes pairs rarely exceed 0.01, probabilities as low as 0.10 could easily be used to discriminate linked from unlinked crimes.

(f) Logistic regression methods

A variety of methods exist for carrying out logistic regression analysis. Two of these methods will be used in the present research. The first method is direct logistic regression, where predictor variables can be entered into the regression model simultaneously (Tabachnik & Fiddell, 1996). This method will be used to examine the predictive accuracy of the various across-crime similarity scores separately.

The second method is forward stepwise logistic regression, where predictor variables can be entered into the regression model in a stepwise fashion (Tabachnik & Fiddell, 1996). As Getty, Seltzer, Tempany, Pickett, Swets and McNeil (1997) 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 will be used to identify the optimal combination of similarity scores for each burglary sample in the present research.

(g) Potential problems with using logistic regression analysis

In relation to the use of logistic regression analysis in the present research, three potential problems must be considered. The first relates to a potential bias that might occur when using logistic regression analysis if very prolific offenders are included in the analysis. If prolific offenders are included in an analysis with non-prolific offenders, and the prolific offenders exhibit high levels of consistency across their crimes, it is possible they will favourably bias the results. That is, it may appear from the results that the majority of offenders in a particular sample are exhibiting a high degree of consistency when in fact the observed levels are only due to one or two offenders. By way of simulation, accurate logistic regression models can be (and have been) created even though all but one offender shows random behaviour across their crimes, if that lone offender shows extremely high levels of consistency.

There is no easy way to control for this potential source of bias other than to make sure that very prolific offenders are not included in any of the samples. As a way of dealing with this in the present research, a uniform distribution of offences will always be selected from each burglary sample for the purpose of constructing regression models. While this has a negative consequence that many offences available for analysis must be ignored, thus raising issues of how generalisable the results will be, this must be viewed as more favourable than the alternative – basing all conclusions on potentially biased results. The problem of generalisability can also be avoided to some extent by ensuring, post analysis, that the results can be applied to the larger samples.

The second potential problem has to due with the validity of the regression models. Even if an effective logistic regression model is constructed on a particular sample of crimes, there is no guarantee that the model will work effectively when applied to other crimes (Efron, 1982; Gong, 1994). This is the case whenever regression models are constructed on one sample but the goal is to apply them to other samples that may differ in a variety of ways (e.g., in terms of when the crimes take place, where the crimes take place, how the crimes take place, etc.).

As a way of dealing with this potential problem, a procedure known as cross-validation will routinely be carried out in the upcoming chapters. The goal of cross-validation in the present case is to ensure adequate model generalisation to crimes not used in the initial model development phase. In the current context, it is a way of increasing ones confidence that the regression models will be applicable to other crimes being committed in the police jurisdiction of interest. For the purpose of validating the logistic regression models developed in the present research, the burglary samples will always be randomly split in half to form an experimental and test sample. The logistic regression models will always be developed on the subset of crime pairs defined as the experimental sample and tested for generalisation on the subset of crime pairs defined as the test sample.

The third potential problem is related to the criterion variable used throughout the present research. Typically, the criterion variable used in regression analysis is statistically independent, in the sense that error associated with one observation is not associated with error from any other observation (Lewis-Beck, 1980). 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 observations include the same offender (e.g., pair 1-1, 1-2 and pair 1-1, 2-3). If the criterion variable is not independent, problems can potentially arise. In such cases, the estimates of standard error corresponding to the regression coefficients, as well as the confidence intervals, can appear smaller than they should be. However, the regression coefficients themselves will not be biased.

This is problematic for two reasons. First, it means that the regression results will indicate we should have higher levels of confidence in the regression coefficients than we actually should have (i.e., the confidence intervals will be too narrow). Second, it means that inferential tests that depend on these estimates of error cannot be relied upon (Chatterjee & Price, 1977). For example, while goodness-of-fit tests will not be problematic, tests used to measure the predictive accuracy of specific predictor variables might be (e.g., Wald's statistic)[5].

In the present research, this potential problem of independence is avoided to a large extent. This is because the measures of predictive accuracy for each type of similarity score, or combination of similarity scores, are generated from ROC analysis rather than from logistic regression analysis (Bennell & Canter, in press). 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 criterion variable is not statistically independent.

3.3.3. Conducting ROC analysis

Once logistic regression models have been constructed on an experimental sample, they are used to calculate estimated probabilities of crimes being linked in the corresponding test sample. These estimated probabilities can then be used to construct empirical ROC curves. As a result of carrying out this procedure, the empirical ROC curves presented throughout this research can be referred to as cross-validated ROC curves. They indicate the level of consistency and discrimination that is expected to be found when analysing solved serial burglary cases that have not been included in the present research.

Having said this, it should be pointed out that the degree of model validity depends on how closely the test samples approximate reality. The solved serial burglaries examined in the present research are probably similar to a portion of

---

[5]As seen in the formula for calculating Wald's, $W=\beta_1/SE(\beta_1)$, this statistic does take into account the estimates of standard error associated with each regression coefficient.

burglaries that will occur some time in the future within each police force. Consequently, it is appropriate for these offences to form part of the test samples. However, more realistic test samples would also have included non-serial burglaries as well as unsolved burglaries if this were possible. Since non-serial burglaries are not included in any of the test samples, the results in the present research should be interpreted cautiously.

It should also go without saying that the generalisability of the logistic regression models across time and across geographic regions should not be taken as given. Indeed, the results presented in upcoming chapters indicate that the regression models may not be generalisable across geographic regions within the same police jurisdiction, let alone across regions located at opposite ends of the country.

(a) The ROC program

All ROC analyses in the present research will be carried out using the ROC sub-routine found in the Statistical Package for the Social Sciences (SPSS). A range of analytical packages now exists for carrying out ROC analysis (Greiner *et al.*, 2000). The ROC routine in SPSS was chosen because it was readily accessible, very easy to use, and capable of dealing with the large numbers of crimes pairs resulting from each of the burglary samples. Having said that, the ROC results presented throughout this research were often confirmed using another ROC program (ROCKIT: Metz, Hermann & Shen, 1988). This was done for two reasons. The first reason was to make sure the ROC results generated using SPSS were consistent with the results that emerged from another commonly used ROC program. The second reason was to get information about the ROC graphs that SPSS could not provide (e.g., an indication of significant differences between ROC curves).

All ROC programs work in essentially the same way. Most start with the continuous output of some decision-maker or analytical procedure. In the present case, this output corresponds to the predicted probabilities that result from running logistic regression analysis on data from a burglary sample. The program then sets a number of cut-off points along this output at intervals related to the total number of different predicted values. Specifically, the larger the number of different values in the output, the narrower the intervals that are set. Values of pH and pFA are then calculated for outputs that exceed each successive cut-off point, and the resulting ROC coordinates are plotted on a ROC graph. Various AUC measures are then automatically produced using one or all of the methods discussed in Appendix B.

(b) Obtaining separate measures of consistency and discrimination

The ROC procedure is used for three general purposes in the present research. The first purpose is to cross-validate the logistic regression models constructed from data in the experimental samples, which will be done by constructing ROC curves based on the predicted probabilities calculated from the test samples. The second purpose is to examine behavioural consistency and overall discrimination, which will be done by examining the AUCs associated with each ROC curve. The third purpose is to examine threshold-specific levels discrimination, which will be done by examining the proportions of each decision outcome resulting at various decision thresholds falling along a ROC curve.

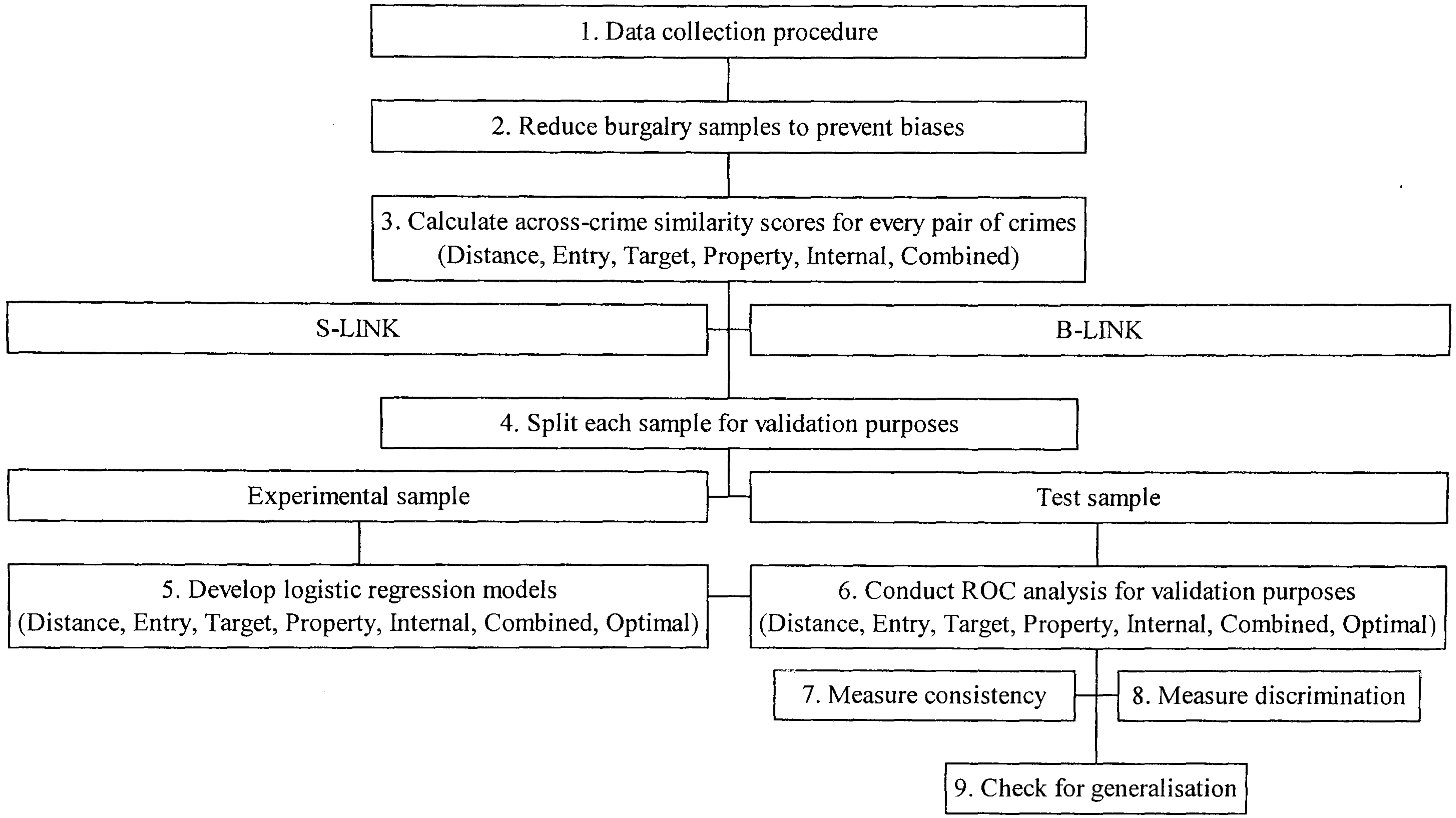
(c) Generalising to larger samples

As stated earlier in this chapter, in order to avoid favourably biasing observed levels of consistency and discrimination by including prolific offenders in any of the samples, a uniform distribution of offences across offenders will be selected from each burglary sample for the purpose of analysis. While this is appropriate, it would still be worthwhile to see if the ROC results generalise to the larger, original samples. At the end of each chapter, random samples of crime pairs will be extracted from the larger samples to examine the degree to which the ROC results generalise. If the ROC results do in fact provide a valid representation of how accurately discriminations can be made, it should be possible to use pre-specified decision thresholds to achieve similar levels of discrimination accuracy across samples of varying sizes.

3.3.4. A summary of the analytical procedure

In order for the analytical procedure used throughout the present research to be totally clear, Figure 3.3 provides a schematic diagram summarising the major steps. The numbers in the boxes refer to the sequence in which the various analytical stages take place. Thus, the process can be seen to consist of: (1) data collection, (2) sample size reduction to prevent biases, (3) calculation of across-crime similarity scores, (4) construction of validation samples, (5) development of logistic regression models, (6) validation of the logistic regression models using ROC analysis, (7) measurement of consistency, (8) measurement of discrimination, and (9) check for generalisation to larger samples.

Figure 3.3. A schematic diagram of the analytical procedure



CHAPTER 4
THE BEHAVIOUR OF SERIAL BURGLARS IN LONDON

4.1. Introduction

In this chapter, the ROC procedure will be formally introduced by exploring the spatial behaviour of residential burglars in London. Specifically, the locations where these offenders commit their crimes, relative to one another, will be investigated. Two research questions in particular will be examined. The first question examines whether crime site selection is expressed in a consistent fashion by residential burglars in London. The second question examines whether an analysis of these choices forms a reliable basis for distinguishing between crimes committed by different London burglars.

Essentially, the answer to the second question depends on the answer to the first. In terms of choosing a crime site, the degree of consistency exhibited by residential burglars in London depends on the extent to which these burglars exhibit high levels of behavioural similarity and distinctiveness. In other words, consistency depends on the degree to which burglars target similar geographic locations across their own crimes and different geographic locations compared to other serial burglars. Numerically, behavioural consistency will exist if the distances between crimes committed by the same offender are small, but the distances between crimes committed by different offenders are large. If relatively high levels of consistency do emerge in this sample of residential burglars, it should be possible to achieve relatively high levels of discrimination accuracy using their inter-crime distances.

Introducing the ROC procedure in this way makes sense for a variety of reasons. First, the police should be able to record the location of crime sites in a more reliable and accurate fashion compared to any other crime scene behaviour. Therefore, unlike other potential forms of behavioural consistency, spatial consistency should not be hidden by problems with police data. Second, there is already evidence suggesting that many offenders exhibit other forms of limited spatial mobility when committing their crimes. For example, the majority of burglars appear to travel very short distances from home to commit at least some of their crimes (Barker, 2000; Baldwin & Bottoms, 1976; Brantingham & Brantingham, 1981; Rengert & Wasilchick, 2000; Wiles & Costello, 2000). Thus, if high levels of consistency and discrimination are going to be found in the behaviours exhibited by residential burglars in London, it seems likely they will be found in relation to some aspect of their spatial behaviour.

4.1.1. The area

The burglaries included in the London sample were committed between April 1999 and April 2000. London covers an area of approximately 336 km^2, has a population of about 7.2 million persons, and a population density of 4611 persons/km^2. London is classified as heavily urbanised and it represents the commercial, financial and political centre of the UK. The Metropolitan Police Service that serves London consists of 25550 police officers. These officers patrol 33 different police districts that correspond to London's borough councils. These boroughs differ drastically from one another in terms of their demographic features and topography making it very difficult to provide a general description of London.

The frequency of reported residential burglary in London, and the associated clear-up rate, varies dramatically across the 33 boroughs. In 2001, for example, the frequency of reported burglary varied from as low as 774 in Sutton to as high as 4748 in Lambeth. Clear-up rates also varied from as low as 6% in Lambeth to as high as 28% in Sutton. In terms of reported crime, residential burglary is not the most serious problem in any of the London boroughs. Compared with residential burglary, for example, rates for criminal damage, theft from motor vehicles, fraud and violence against persons are often much higher. However, the high volume of residential burglaries, coupled with the relatively low clear-up rate, does suggest that residential burglaries represent a serious problem for the police in London.

4.1.2. The data

The data in this sample consists of x and y geo-coded coordinates indicating the position of burglary locations across the boroughs to the nearest metre. The entire London sample includes 69 serial burglars responsible for a total of 816 solved residential burglaries. The offence series range in length from 5 crimes to 37 crimes. For the purpose of the analysis presented in this chapter, however, a smaller subset of crimes was selected from this large sample. This smaller sample will henceforth be referred to as the reduced London burglary sample. Specifically, 5 crimes from each offender were randomly selected from the entire London sample to form a reduced London sample of 345 burglaries.

As discussed in Chapter 3, this procedure controls for any bias that may be introduced into the analysis as a result of sampling from an uneven distribution of offences across offenders. For example, if the offender who committed the 37 crimes exhibited an extremely high level of consistency over time, this could significantly bias the results and lead to conclusions that behaviour was consistently expressed by the majority of offenders when, in fact, it was not. However, this selection process also brings with it a number of negative consequences, the primary one being a reduction in how representative the results may be of residential burglaries committed in London. While this is a serious concern, the problem will be avoided to a degree by ensuring the results do in fact generalise to the larger sample.

4.2. Calculating spatial similarity scores

In order to examine this aspect of spatial behaviour for residential burglars in London, spatial similarity scores needed to be calculated. To calculate these scores, the x and y coordinates from each of the burglaries in the reduced London sample were entered into *S-LINK*. As mentioned already, this computer program calculates, for every possible crime pair in the sample, the distance in kilometres between each and every crime. All of the generated crime pairs are then defined as unlinked or linked, based on who is known to have committed the crimes. This resulted in spatial similarity scores for 58649 unlinked and 691 linked crime pairs. These similarity scores form the basis of all analyses in this chapter.

4.3. A descriptive analysis of the spatial similarity scores

Before moving on to a more in-depth analysis of the data, a descriptive analysis of the spatial similarity scores was carried out. Descriptive statistics were calculated across all unlinked and linked crime pairs in the reduced London sample. These values were then compared in order to determine if they were significantly different. Recall the general hypothesis being tested throughout the present research is that offenders exhibit statistically demonstrable levels of consistency in their criminal behaviour. Thus, it is expected that similarity scores calculated across unlinked crime pairs will be low relative to the similarity scores calculated across linked crime pairs. If this were found to be the case, it would provide support that this aspect of spatial behaviour is expressed consistently over time, thus making it possible to use this information to discriminate between crimes committed by different offenders.

The descriptive analysis of the spatial similarity scores is presented in Table 4.1. This table includes the mean values of the similarity scores, along with their ranges and standard deviations. In addition, the result from a t-test is provided. As indicated by the mean values, unlinked crime pairs have lower levels of spatial similarity compared to linked crime pairs. The average inter-crime distance for unlinked crime pairs is 15.70 km whereas the average inter-crime distance for linked crime pairs is 1.80 km. In other words, crimes committed by different offenders are more geographically dispersed compared to crimes committed by the same offender.

Despite the fact that unlinked crimes are more geographically dispersed, it is clear from the range of similarity scores observed that cases exist where high levels of spatial similarity are found for unlinked crimes, as well as low levels of similarity for linked crime pairs. This suggests that while there is a tendency for unlinked crime pairs to be more geographically dispersed, the findings reflect merely that, a tendency. Nevertheless, a highly significant difference is found between the spatial similarity scores, as indicated by the results from the t-test ($t=47.53$, $df=59338$, $p \leq 0.001$).

Table 4.1. Summary of the London residential burglary data

	Unlinked crime pairs (n=58649)			Linked crime pairs (n=691)			
Variable	M	Range	SD	M	Range	SD	t
Distance	15.70	0-40.31	0.77	1.80	0-25.30	0.30	47.53***

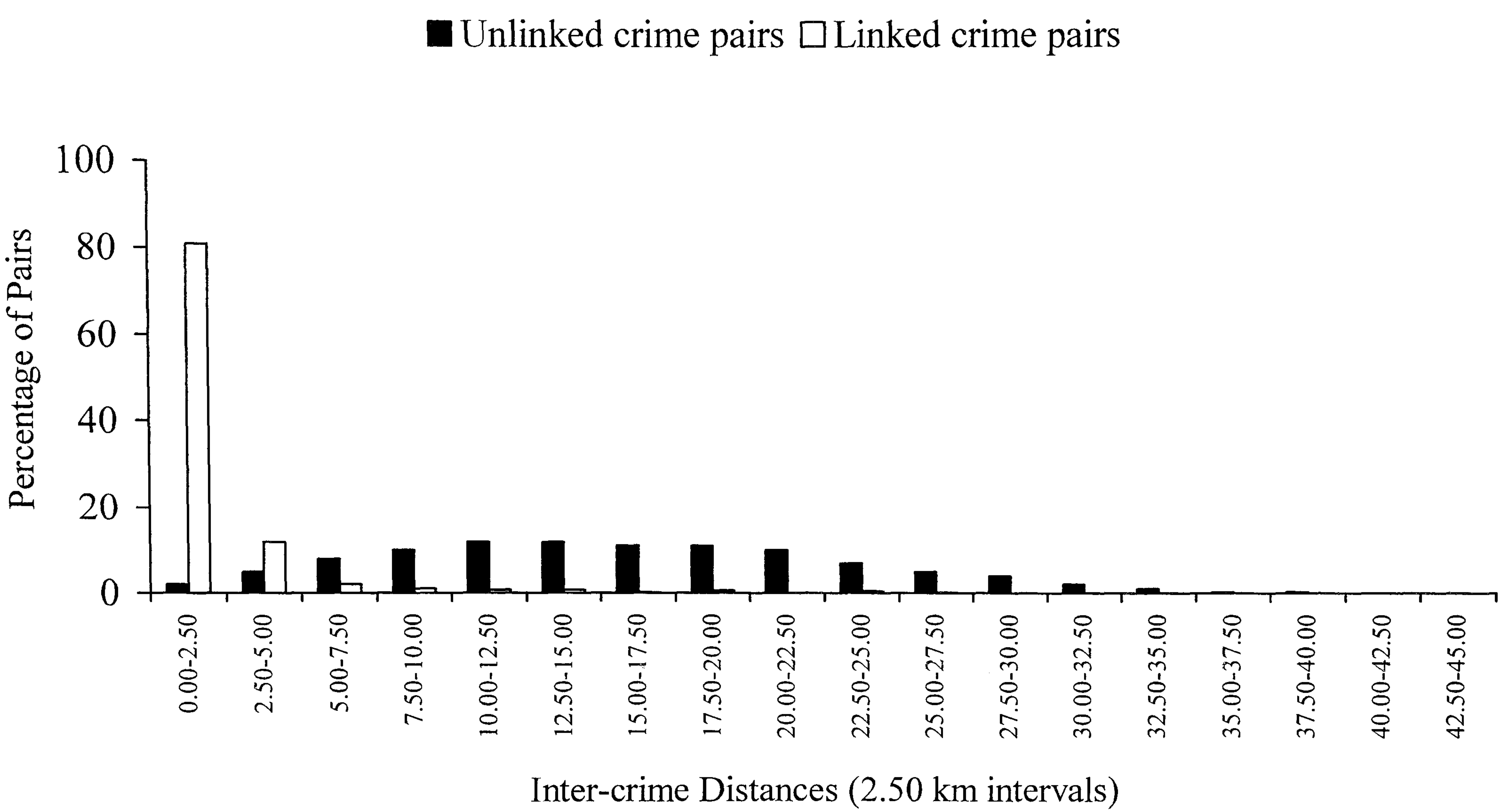
*: $p \leq 0.05$; **: $p \leq 0.001$; ***: $p \leq 0.001$

4.3.1. The distribution of spatial similarity scores

The degree to which unlinked and linked crime pairs differ in terms of their inter-crime distances can be made even clearer by examining the distribution of spatial similarity scores. This is done for the reduced London sample in Figure 4.1 using distance intervals of 2.50 km. What this figure clearly indicates is that the vast majority of crimes committed by the same offender in London are separated by relatively small inter-crime distances. Indeed, about 80% of all linked crime pairs are characterised by inter-crime distances shorter than 2.50 km. The distances between crimes committed by different offenders, on the other hand, are more evenly distributed across the distance intervals.

These sorts of distributions strongly suggest that it should be possible to accurately discriminate between unlinked and linked crime pairs. Unfortunately, the fact that there is a degree of overlap between unlinked and linked crime pairs with respect to their inter-crime distances also suggests that the level of spatial consistency exhibited by burglars in London will not be absolute and that, as a result, discrimination accuracy will not be perfect. Essentially, what the distributions of spatial similarity scores indicate is that regardless of where a decision threshold is placed along the x-axis in Figure 4.1 a number of discrimination errors should be expected. The fact that the degree of overlap is relatively small, however, suggests that the number of errors will probably be marginal.

Figure 4.1. Distribution of inter-crime distances



4.4. Logistic regression analysis

The descriptive analysis presented in the previous section provides some initial support that residential burglars in London choose their crime site locations in a relatively consistent fashion. This, in turn, suggests it may be possible to use the relative distances between burglary locations to discriminate between crimes committed by different offenders. In order to move beyond a simple examination of average inter-crime distances to a stage where consistency and discrimination can be accurately measured, a logistic regression model was constructed. Beyond providing further evidence of the degree to which burglaries in London can be discriminated from one another, this model provides the necessary data for ROC analysis. It is only by using ROC analysis that accurate measures of consistency and discrimination can be calculated.

4.4.1. A validation dataset

For the purpose of validating the logistic regression model, the dataset containing all the crime pairs generated from the reduced London sample were split randomly in half to form an experimental and test sample. The logistic regression model was developed on the experimental sample and tested for generalisation on the test sample. Table 4.2 contains a summary of the logistic regression model constructed from the experimental sample. A range of information is provided in this table, including the model coefficients and standard errors (constant and logit), an indicator of predictive accuracy (Wald's statistic), and indices of general model fit (R^2 and X^2).

Table 4.2. Logistic regression model for London residential burglary data [a,b]

Model[c]	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-0.04 (0.10)	-0.83 (0.04)	503.39 (1)***	0.53	1919.07 (1)***

[a] Criterion variable (unlinked crime pair=0, linked crime pair=1); [b] Sample size =59340 crime pairs (58649 unlinked, 691 linked); [c] Direct logistic regression analysis was used to construct this model; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

4.4.2. The regression coefficients

Consistent with the findings presented in Table 4.1, the negative sign of the logit coefficient in the logistic regression model indicates that unlinked crime pairs tend to be characterised by larger inter-crime distances than linked crime pairs (logit=-0.83). To determine what this logit coefficient actually means it was exponentiated. Considering that the average inter-crime distance for linked crime pairs in the reduced London sample is 1.80 km, it makes sense to examine the effect of increasing inter-crime distances using intervals of 0.10 km. This can be done by multiplying the logit coefficient in Table 4.2 by 0.10 and exponentiating it, as in $e^{(0.10 \times -0.83)}=0.92$. That is, for every increase of 0.10 km between any two crimes in the reduced London sample, the odds that the crimes are linked are multiplied by 0.92, which would reduce them. This is consistent with the fact that unlinked crime pairs in London are characterised by larger inter-crime distances than linked crime pairs.

The impact of changes in inter-crime distances was also assessed by examining the probability of two crimes being linked in the reduced London sample. Recall that this probability can be estimated by calculating the logit, transforming the logit into odds, and converting the odds into a probability. Thus, given the model in Table 4.2, and two crimes that are 0.10 km apart, the logit can be calculated, -0.04-0.83(0.10)=-0.12. The logit can then be transformed into odds by exponentiating it, $e^{-0.12}=0.88$. Finally, the odds can be converted into *p*(linked) by dividing it by 1 plus the odds, 0.88/(1+0.88)=0.47. Thus, the probability of two crimes being linked when they are 0.10 km apart is p=0.47. These probabilities can then be compared to the probability of two crimes being linked when they are 0.20 km apart (p=0.44), 0.30 km apart (p=0.43), and so on. As is evident from these few calculations, the general pattern in this sample is for p(linked) to decrease slightly as inter-crime distances increase.

4.4.3. Predictive accuracy and goodness-of-fit

Also consistent with the findings in Table 4.1, the logistic regression model is found to have a high level of predictive accuracy and a high degree of fit with the data. The model's predictive accuracy is indicated by a highly significant Wald's statistic (W=503.39, df=1, $p \leq 0.001$). However, as pointed out in Chapter 3, this value must be treated with an appropriate level of caution due to its reliance on the standard error of the logit coefficient. The model's fit with the data is indicated by a high R^2 value (R^2=0.53), which indicates the proportion of variance in the criterion variable explained by the regression model, and a highly significant X^2 value (X^2=1919.07, df=1, $p \leq 0.001$).

4.5. ROC analysis

In the previous section, a logistic regression model constructed from data in the experimental sample indicated that the distance between crime locations could be used with a relatively high degree of accuracy when predicting whether crime pairs are unlinked or linked. The output from this regression analysis will now be used to create an empirical ROC graph. This will make it possible to validate the regression model and to obtain separate and accurate measures of both consistency and discrimination.

4.5.1. Transforming frequencies into proportions

As already stated, ROC analysis involves calculating the proportion of decision outcomes made in a two-alternative decision task across different decision thresholds. As a formal introduction to the procedure, this is illustrated in Table 4.3 for 9 decision thresholds using the logistic regression model presented above. The first step in calculating these proportions is to use the logistic regression model to estimate, for every pair of crimes in the test sample, the probability that each pair is linked. The next step is to set different decision thresholds along this continuous output, whereby any crime pair having a probability exceeding the threshold is classified as linked. At that point, it is possible to calculate the frequency of decision outcomes resulting for each decision threshold. Each of these frequencies can then be converted into a conditional probability at each decision threshold. The conditional probabilities of each decision outcome are estimated from their respective frequencies using the following four formulae: pH=H/(H+M), pM=M/(H+M), pCR=CR/(CR+FA), pFA=FA/(CR+FA). See Appendix B for a more thorough discussion of these calculations.

Table 4.3. Converting frequency data into proportions for ROC analysis

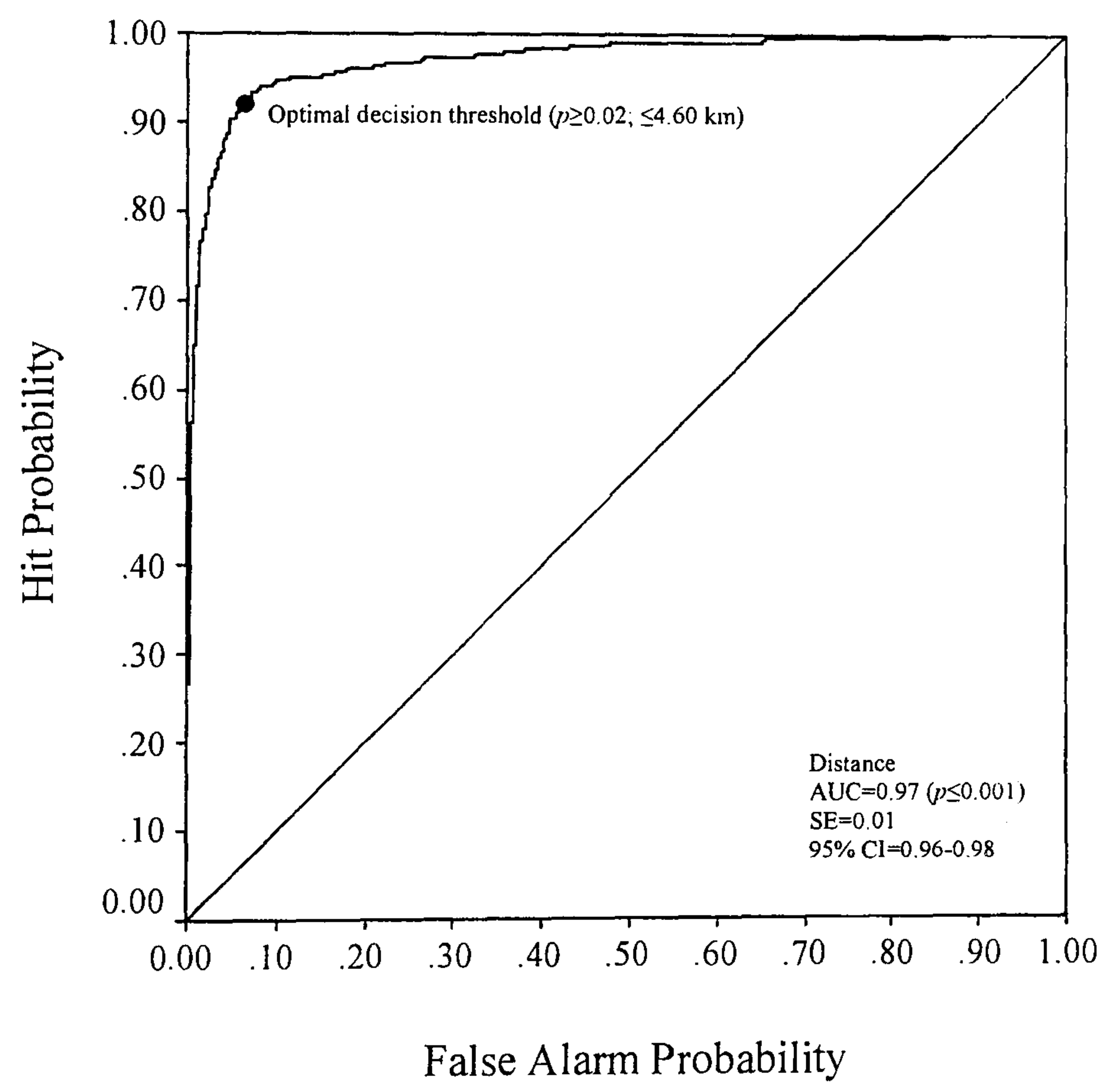
	Frequencies				Proportions			
Threshold (distance)	H	M	CR	FA	*p*H	*p*M	*p*CR	*p*FA
$p \geq 0.01$ ($\leq$5.20 km)	324	21	26648	2676	0.94	0.06	0.91	0.09
$p \geq 0.02$ ($\leq$4.60 km)	320	25	27316	2008	0.93	0.07	0.93	0.07
$p \geq 0.03$ ($\leq$4.00 km)	315	30	27658	1666	0.91	0.09	0.94	0.06
$p \geq 0.04$ ($\leq$3.70 km)	312	33	27874	1450	0.90	0.10	0.95	0.05
$p \geq 0.05$ ($\leq$3.40 km)	305	40	28038	1286	0.88	0.12	0.96	0.04
$p \geq 0.06$ ($\leq$3.20 km)	299	46	28169	1155	0.87	0.13	0.96	0.04
$p \geq 0.07$ ($\leq$3.00 km)	298	47	28282	1042	0.86	0.14	0.96	0.04
$p \geq 0.08$ ($\leq$2.80 km)	292	53	28398	926	0.85	0.15	0.97	0.03
$p \geq 0.09$ ($\leq$2.70 km)	289	56	28490	834	0.84	0.16	0.97	0.03

H: the frequency of hits; M: the frequency of misses; CR: the frequency of correct rejections; FA: the frequency of false alarms; *p*H: the conditional probability of hits; *p*M: the conditional probability of misses; *p*CR: the conditional probability of correct rejections; *p*FA: the conditional probability of false alarms

4.5.2. Developing an empirical ROC curve

Once these conditional probabilities have been calculated, they are plotted on a graph as a function of the different thresholds to form an empirical ROC curve. The empirical ROC curve calculated using data from the test sample, along with its AUC (and *p*-value), standard error, and 95% confidence interval, is presented in Figure 4.2. The data plotted in this ROC graph is the same as that presented in Table 4.3, except many more decision thresholds have been used. As discussed in Chapter 2, the resulting ROC curve can provide useful information about a variety of issues. First, the ROC curve can be used to indicate the extent to which offenders are consistent across the crimes they commit with respect to certain aspects of offence behaviour. This can be done by examining the AUC associated with the ROC curve. Second, the ROC curve can be used to indicate threshold-specific levels of discrimination accuracy. This can be done by examining the operating points falling along the ROC curve.

Figure 4.2. ROC graph for London residential burglary data



4.5.3. The AUC as a measure of spatial consistency

Recall that the height of the ROC curve corresponds to the degree of overlap between the spatial similarity scores calculated across unlinked and linked crime pairs. As a result, the AUC can be used as a measure of spatial consistency. Specifically, the higher the ROC curve is in its graph, the smaller the amount of overlap between the two probability distributions of similarity scores (i.e., the higher the proportion of hits to false alarms at any given decision threshold).

Thus, a high ROC curve indicates the behavioural feature used to derive the curve is expressed in a consistent fashion, while a low ROC curve indicates lower levels of consistency.

Consistent with the analysis of data in the experimental sample, the ROC curve in Figure 4.2 suggests that the crime site choices made by residential burglars in London are an extremely consistent feature of their criminal behaviour. Indeed, the AUC in this case is equal to 0.97, which is remarkably high and significantly greater than chance ($p \leq 0.001$). This AUC indicates that there is little overlap between the spatial similarity scores calculated across unlinked and linked crime pairs within the reduced London burglary sample.

4.5.4. Operating points as a measure of spatial discrimination

While the AUC corresponds to an overall measure of consistency in crime site selection, and an overall measure of discrimination accuracy, it is also possible to identify threshold-specific measures of discrimination accuracy using the ROC curve. As illustrated initially in Figure 4.1 and later in Table 4.3, where the decision threshold is placed can have a serious impact on discrimination accuracy. Consequently, it is essential to choose an appropriate decision threshold for discrimination purposes.

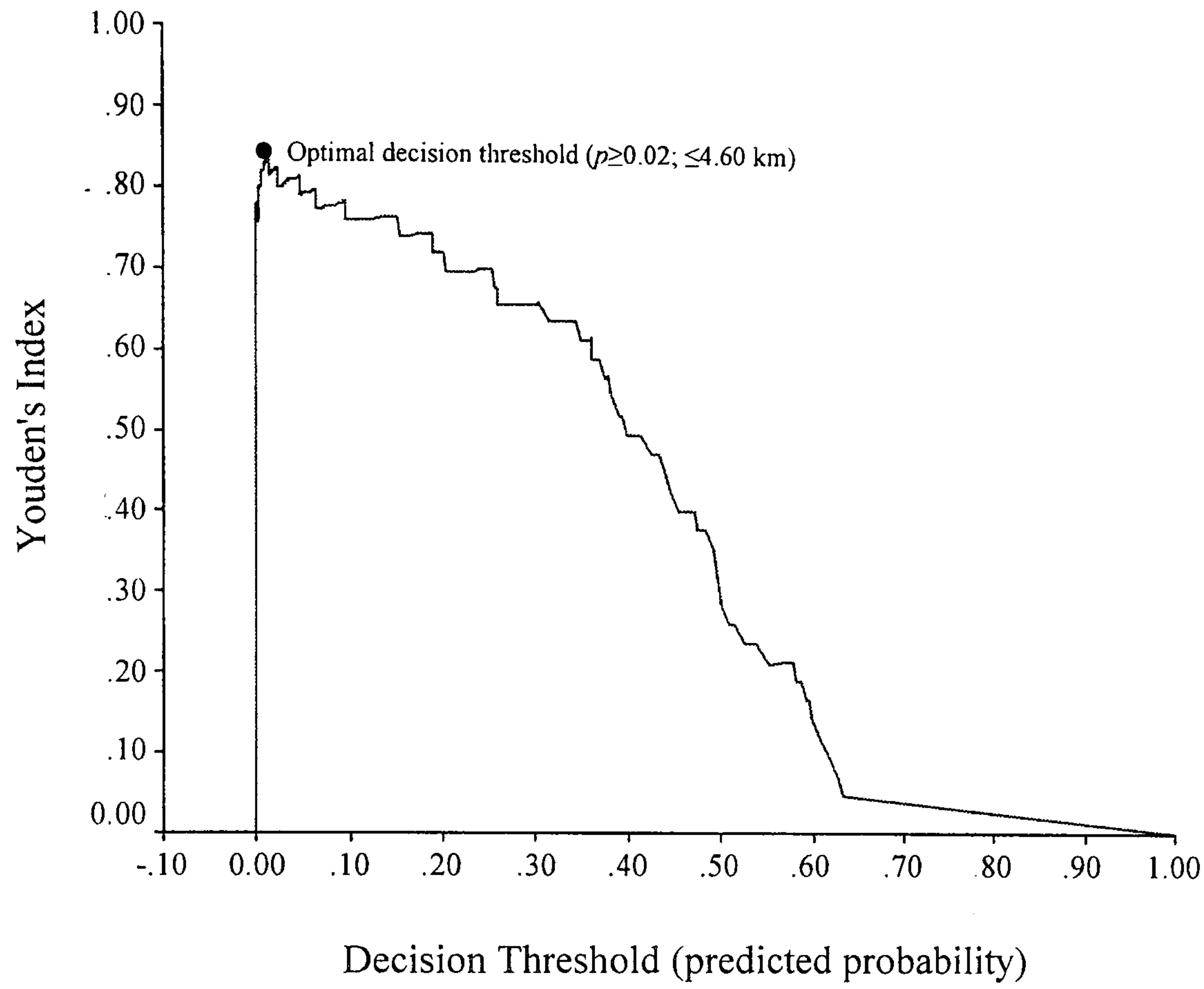
As discussed more thoroughly in Chapter 2, there are a variety of standard ways in which appropriate decision thresholds can be set. Again, the most effective procedure is to take into account the prevalence of the two alternatives under consideration as well as all the costs and benefits associated with each decision outcome. However, to use this type of approach in the present research would be extremely difficult without an in-depth examination of the economic, as well as the ethical, costs and benefits associated with incorrect and correct linking decisions.

As an alternative, the decision threshold that results in the maximum number of hits and the minimum number of false alarms was identified. This can be done formally using the maximisation formula presented in Chapter 3, which is also referred to as Youden's index (Hilden, 1991). Youden's index is represented simply as J=pH+pCR-1, where the subtraction of 1 from pH + pCR ensures that J always lies between 0 and 1. The goal is to choose the decision threshold that results in the highest possible J value, where both pH and pCR are equal to 1 (i.e., no incorrect linking decisions are made).

As illustrated in Figure 4.3, this occurs in the present case at an approximate decision threshold of $p \geq 0.02$ ($\leq$4.60 km). At this particular threshold, pH=0.93 (pM=0.07), pCR=0.93 (pFA=0.07), and J=0.84. Looking back at Figure 4.2, it can be seen that this decision threshold falls at a point on the ROC curve closest to the upper left corner of the graph, as would be expected. Furthermore, looking back at Figure 4.1, it can be seen that this decision threshold falls at a point along the x-axis that should accurately discriminate between unlinked and linked crime pairs.

One of the reasons the optimal decision threshold (i.e., $p \geq 0.02$) is so low in this case is because of the large discrepancy between the number of unlinked and linked crime pairs occurring in the reduced London sample. As indicated in Appendix A, predicted probabilities, such as those produced by logistic regression analysis, incorporate prior probabilities. Prior probabilities refer to the probability that an unlinked or linked crime pair will actually exist. Since the prior probability of a linked crime pair existing in the London sample is extremely low, the predicted probability of two crimes being linked will also be low. This will be the case even when the same offender has in fact committed the crimes. Consequently, the optimal decision threshold for the reduced London sample must also be low.

Figure 4.3. Identifying an appropriate threshold using Youden's index



4.6. Validating the empirical ROC curve

When the logistic regression model developed using data from the experimental sample is applied to data from the test sample, the resulting ROC curve is extremely high in its graph. This suggests that the logistic regression model does generalise to crimes beyond those used to construct the original regression model. The validity of the model can be tested more directly, however, by creating a ROC curve based on the original experimental sample. When this is done, the ROC curve that results also has an AUC of 0.97 indicating a high degree of validity across the two samples.

4.6.1. External discrimination trials

Another form of validation would also be useful to confirm that the information contained within a ROC graph transfers to discrimination tasks as the police usually conduct them. To examine this issue, a number of external discrimination trials were carried out, which involved extracting random samples of crime pairs from the reduced and the entire London sample. The objective is to determine whether the values of pH and pFA associated with the optimal decision threshold in Figure 4.2 correspond to the values of pH and pFA obtained across discrimination trials when the optimal threshold is used to make the predictions.

Specifically, a value 1 was placed next to a crime pair in the extracted samples every time the inter-crime distance was $\leq$4.60 km, otherwise a value of 0 was placed next to the pair. The accuracy of these predictions was then measured directly by forming two-way contingency tables of predictions versus reality. From these tables pH, pM, pCR and pFA were calculated and related back to the ROC graph to see how they compared. Recall that pH and pFA at the optimal threshold equal 0.93 and 0.07 respectively. Therefore, if the ROC graph provides a valid representation of how accurately real discriminations can be made it should be possible to achieve similar levels of accuracy across the random trials when the same decision threshold is used.

The results of these trials are presented in Table 4.4. Across 5 randomly selected samples of 10000 crime pairs, the average hit and false alarm rates were 0.95 and 0.07 respectively. Across 5 randomly selected samples of 50000 crime pairs, the average hit and false alarm rates were 0.87 and 0.07 respectively. These rates closely correspond to the predicted values regardless of the sample size. Furthermore, on every trial there is a highly significant association between predictions and reality, as reflected by the X^2 value presented in the last column of the table. This indicates that if a pair of crimes included in the discrimination trials was, in reality, a linked crime pair, the prediction was also more likely to be that the crimes were linked.

4.7. Chapter summary

In this chapter, the spatial behaviour of residential burglars in London was explored. Specifically, the distances between crime sites chosen by the same or different burglars were examined. This was done as a way of determining whether burglars choose their crime sites in a consistent fashion, and if so, whether an analysis of the distances between crime sites could provide a reliable basis for distinguishing between crimes committed by different offenders. An initial descriptive analysis of spatial similarity scores provided strong support for this hypothesis. A highly significant difference was found between the inter-crime distances for unlinked and linked crimes pairs. As expected, the burglaries committed by different offenders tended to be more geographically dispersed.

In order to move beyond an examination of average inter-crime distances, logistic regression analysis and ROC analysis were used. Logistic regression analysis provided predicted probabilities that crimes were linked and it was these probabilities that formed the basis of the ROC analysis. The ROC results confirmed that crime sites are selected in a consistent fashion by burglars in London. Indeed, the AUC associated with inter-crime distances was 0.97. In line with this, high levels of discrimination accuracy could also be achieved. Using a threshold of ≤4.60 km, 93% of linked burglaries were correctly classified as linked and only 7% of unlinked burglaries were incorrectly classified as linked. Similar results emerged across discrimination trials of various sizes, suggesting that the ROC procedure has an adequate level of predictive accuracy in London.

From a theoretical perspective, this finding extends existing research, which suggests that burglars are usually very limited in terms of their spatial mobility (e.g., Brantingham & Brantingham, 1981). While the vast majority of this research has focused on journey to crime distances, the results presented in this chapter clearly indicate that residential burglars in London are spatially limited in another way as well. Specifically, residential burglars in London establish offending territories that do not overlap to any great extent with territories established by other burglars. From a practical perspective, this finding contributes to the small amount of research that has attempted to develop methods for linking serial burglaries. Until now, the approach has been based primarily on an analysis of crime scene behaviours (e.g., Green *et al.*, 1976). However, the results presented in this chapter suggest that the analysis of a burglar's spatial behaviour may be able to significantly enhance discrimination performance.

Table 4.4. Validation trials for London residential burglary data

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.02$ (≤ 0.46 km)	10000	0.96 (121)	0.04 (5)	0.93 (9195)	0.07 (679)	1343.63 (1)***
	$p \geq 0.02$ (≤ 0.46 km)	50000	0.88 (353)	0.12 (50)	0.93 (46153)	0.07 (3444)	3705.76 (1)***
2	$p \geq 0.02$ (≤ 0.46 km)	10000	0.95 (101)	0.05 (5)	0.93 (9238)	0.07 (656)	1178.02 (1)***
	$p \geq 0.02$ (≤ 0.46 km)	50000	0.88 (337)	0.12 (48)	0.93 (46200)	0.07 (3415)	3580.11 (1)***
3	$p \geq 0.02$ (≤ 0.46 km)	10000	0.98 (117)	0.02 (3)	0.93 (9226)	0.07 (654)	1376.17 (1)***
	$p \geq 0.02$ (≤ 0.46 km)	50000	0.87 (340)	0.13 (52)	0.93 (45902)	0.07 (3706)	3285.57 (1)***
4	$p \geq 0.02$ (≤ 0.46 km)	10000	0.96 (112)	0.04 (5)	0.93 (9215)	0.07 (676)	1287.07 (1)***
	$p \geq 0.02$ (≤ 0.46 km)	50000	0.83 (322)	0.17 (65)	0.93 (45983)	0.07 (3630)	3038.63 (1)***
5	$p \geq 0.02$ (≤ 0.46 km)	10000	0.91 (99)	0.09 (10)	0.93 (9215)	0.07 (676)	1063.82 (1)***
	$p \geq 0.02$ (≤ 0.46 km)	50000	0.88 (365)	0.12 (50)	0.93 (45988)	0.07 (3597)	3673.89 (1)***
Average	$p \geq 0.02$ (≤ 0.46 km)	10000	0.95 (110.00)	0.05 (5.60)	0.93 (9219.40)	0.07 (665.00)	--
	$p \geq 0.02$ (≤ 0.46 km)	50000	0.87 (343.40)	0.13 (53)	0.93 (46045.20)	0.07 (3558.40)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

CHAPTER 5
THE BEHAVIOUR OF SERIAL BURGLARS IN DORSET

5.1. Introduction

In contrast to the previous chapter, which examined the spatial behaviour of residential burglars in London, this chapter explores the behaviour of residential burglars in Dorset. Also in contrast to the previous chapter, which dealt solely with the spatial aspect of residential burglary behaviour, the data collected from Dorset allows a variety of burglary behaviours to be examined. These additional behavioural domains include entry behaviours, target selection choices, property stolen and internal behaviours. Despite these differences, the primary objectives in this chapter are the same as they were in the previous chapter. The first objective is to determine whether these aspects of residential burglary behaviour are expressed in a consistent fashion by burglars in Dorset. The second objective is to determine whether an analysis of these behaviours can form a reliable basis for distinguishing between crimes committed by different offenders.

As before, the answer to the second question depends on the answer to the first, and both questions can be examined by comparing across-crime similarity scores for unlinked and linked crimes pairs. However, there are a variety of reasons to suspect that the level of consistency and discrimination associated with crime scene behaviours may not be as high as the levels found for inter-crime distances. The first reason is that while inter-crime distances can be recorded in a very reliable and accurate fashion, the same level of reliability and accuracy is not likely to be associated with crime scene behaviours (Alison *et al.*, 2001). The second reason is that while evidence exists that most burglars exhibit limited spatial mobility across the crimes they commit, crime scene behaviours are thought to be influences by various factors (Davies, 1992; Douglas & Munn, 1992, Turvey, 2000).

5.1.1. The area

The burglaries included in the Dorset sample were committed between January 1997 and December 1998 across Dorset. Dorset is a relatively large county located in the southern most part of the UK. It covers an area of 608 km^2, has a population of approximately 688170 persons, and an average population density of about 151 persons/km^2. The county is largely rural with only 6.3% of its entire area classified as urban.

As of January 2001, the Dorset Police Service consisted of 1378 police officers. These officers patrol 4 different police divisions, including Western division, Eastern division, Poole division and Bournemouth division. Each of these divisions consists of numerous geographic areas that vary in many ways making it very difficult to provide a general description of the whole county. As an example, Western division consists of Bridport, Dorchester, Sherborne, East Weymouth and West Weymouth. Across these 5 areas the topography varies from coastline to farmland to villages to commercial districts. The population of these areas also varies from 17500 in Dorchester to 41000 in West Weymouth, as does the density, from 32 persons/km^2 in West Weymouth to 200 persons/km^2 in Dorchester.

The frequencies of reported residential burglaries, and the associated clear-up rates, also vary across the 5 divisions. In 2001, for example, the frequency of residential burglary varied from as low as 427 in Western division to as high as 1028 in Bournemouth division. Clear-up rates also varied from as low as 13% in Eastern division to as high as 22% in Bournemouth division. In terms of reported crime, residential burglary is not the most serious problem in any of the police divisions. Compared with residential burglary, for example, rates for criminal damage, theft from motor vehicles, fraud and violence against persons are often much higher. However, the high volume of residential burglaries across the divisions along with the relatively low clear-up rates, does suggest that residential burglaries represent a serious problem for the police in Dorset.

5.1.2. The data

As with the London data, the information pertaining to spatial behaviour consists of x and y geo-coded coordinates indicating the position of burglary locations to the nearest metre. For all other behaviours, the data is coded in dichotomous form, with a 1 indicating the presence of a particular crime scene behaviour and a

0 indicating its absence. A more detailed list of the behaviours included in the Dorset sample can be found in Appendix E.

The entire Dorset sample consists of 28 serial burglars responsible for 233 crimes. The offence series range in length from 5 crimes to 10 crimes. For the purpose of the analysis presented in this chapter, however, a smaller subset of crimes was selected from this large sample. This smaller sample will henceforth be referred to as the reduced Dorset sample. Specifically, 5 crimes from each offender were randomly selected from the entire sample to form a reduced Dorset sample of 140 crimes. As in the previous chapter, this was done to control for any bias that may be introduced as a result of sampling from an uneven distribution of offences across offenders. However, because this procedure may cause a reduction in how representative the results are, an attempt was made to determine whether the results do generalise to the entire Dorset sample.

5.2. Calculating spatial and behavioural similarity scores

In order to examine the behaviour of residential burglars in Dorset, spatial and behavioural similarity scores were calculated. The spatial similarity scores again consist of inter-crime distances, calculated by entering the x and y coordinates into *S-LINK*. The behavioural similarity scores consist of Jaccard coefficients. These scores were calculated by inputting the dichotomous data for each of the behavioural domains into *B-LINK*. The exception to this was a behavioural component labelled 'combined', which was obtained by collapsing across the dichotomous data for entry behaviours, target selection choices, property stolen and internal behaviours. *B-LINK* computes, for every possible crime pair in the sample, a Jaccard coefficient between each and every crime. This is done for each behavioural domain separately (i.e., entry behaviours, target selection choices, property stolen and internal behaviours) plus the combined component (i.e., entry behaviours *and* target selection choices *and* property stolen *and* internal behaviours).

All of the generated crime pairs from *S-LINK* and *B-LINK* can then be defined as unlinked or linked, based on who is known to have committed the crimes. This procedure resulted in spatial and behavioural similarity scores for 9450 unlinked crime pairs and 280 linked crime pairs. These similarity scores form the basis of all analyses in this chapter.

5.3. A descriptive analysis of the spatial and behavioural similarity scores

As in the previous chapter, before moving on to a more in-depth analysis of the data, a descriptive analysis of the spatial and behavioural similarity scores was carried out. Descriptive statistics were calculated across all unlinked and linked crime pairs. As before, it was expected that similarity scores calculated across unlinked crime pairs would be low relative to the similarity scores calculated across linked crime pairs. If this were found to be the case, it would provide support that residential burglary behaviours in Dorset are expressed consistently over time, thus making it possible to discriminate between crimes committed by different offenders.

The descriptive analysis of the spatial and behavioural similarity scores is presented in Table 5.1. This table includes the mean values of the similarity scores, along with their ranges and standard deviations. In addition, the results from t-tests are provided. As indicated by the mean values, unlinked crime pairs consistently have lower levels of behavioural similarity scores compared to linked crime pairs. This is the case, both in relation to spatial behaviour as well as the other behavioural domains.

Despite the fact that unlinked crimes are characterised by lower similarity scores, cases exist where high levels of similarity are found for unlinked crime pairs, as well as low levels of similarity for linked crime pairs. Indeed, with the exception of the combined component, the range of similarity scores is 0 to 1.00 for both unlinked and linked crime pairs. Nevertheless, significant differences are consistently found between the similarity scores. Each of the t-tests presented in Table 5.1 indicate highly significant differences on the order of $p \leq 0.001$, with the exception of property stolen where a significant difference is found at a level of $p \leq 0.01$.

Table 5.1. Summary of the Dorset residential burglary data

	Unlinked crime pairs (n=9450)			Linked crime pairs (n=280)			
Variables	M	Range	SD	M	Range	SD	t
Distance	24.39	0-72.25	23.37	2.84	0-19.25	3.72	15.42***
Combined	0.22	0-1.00	0.12	0.28	0-0.65	0.13	8.39***
Entry	0.32	0-1.00	0.22	0.41	0-1.00	0.25	6.79***
Target	0.29	0-1.00	0.25	0.35	0-1.00	0.27	4.41***
Internal	0.17	0-1.00	0.29	0.24	0-1.00	0.34	3.85***
Property	0.12	0-1.00	0.19	0.15	0-1.00	0.23	2.55**

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

5.4. Logistic regression analysis

The descriptive analysis presented in the previous section provides support for the hypothesis that burglars in Dorset express a range of offence behaviours in a consistent fashion. This, in turn, suggests it may be possible to use residential burglary behaviours to discriminate between crimes committed by different offenders. As was the case in the previous chapter, however, the fact that there is a degree of overlap between unlinked and linked similarity scores suggests that the level of consistency exhibited by burglars in Dorset will not be absolute, and that discrimination accuracy will not be perfect. Furthermore, the fact that the degree of overlap varies depending on the behavioural domain considered suggests that consistency and discrimination levels will be domain specific.

To examine these issues more directly, logistic regression models were constructed. This takes the analysis of burglary behaviours beyond a simple examination of average across-crime similarity scores. In this case, a regression model was developed for each of the behavioural domains in isolation, and an optimal model was developed consisting of those behaviours that combine to have the highest level of predictive accuracy. Beyond providing further evidence of the degree to which residential burglaries in Dorset can be discriminated from one another, these models provide the necessary data for ROC analysis.

5.4.1. A validation dataset

As was the case in the previous chapter, the dataset containing all the crime pairs generated from the reduced Dorset sample were split randomly in half for the purpose of validating the logistic regression models. The logistic regression models were developed on the experimental sample and tested for generalisation on the test sample. Table 5.2 contains a summary of each logistic regression model including the model coefficients and standard errors (constant and logit), an indicator of predictive accuracy for each predictor variable (Wald's statistic), and indices of general model fit (R^2 and X^2).

Table 5.2. Logistic regression models for Dorset residential burglary data [a,b]

Model[c]	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-1.68 (0.14)	-0.30 (0.04)	71.45 (1)***	0.22	251.91(1)***
Combined	-4.21 (0.19)	2.83 (0.63)	19.92 (1)***	0.02	18.19 (1)***
Entry	-3.94 (0.16)	1.19 (0.36)	11.21 (1)***	0.01	10.65 (1)***
Target	-3.76 (0.14)	0.77 (0.32)	5.64 (1)**	0.01	5.31 (1)*
Property	-3.64 (0.10)	0.86 (0.37)	5.47 (1)*	0.00	4.86 (1)*
Internal	-3.62 (0.10)	0.53 (0.26)	4.05 (1)*	0.00	3.79 (1)*
Optimal	-2.24 (0.20)	--	--	0.24	269.70 (3)***
Distance	--	-0.29 (0.03)	72.33 (1)***	--	--
Entry	--	1.20 (0.36)	11.35 (1)***	--	--
Internal	--	0.69 (0.27)	6.47 (1)**	--	--

[a] Criterion variable (unlinked crime pair=0, linked crime pair=1); [b] Sample size=9730 crime pairs (9450 unlinked, 280 linked); [c] Direct logistic regression analysis was used to construct the single feature models; forward stepwise logistic regression analysis was used to construct the optimal model (inclusion criteria: $p \leq 0.05$); *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

5.5. The single feature models

Each of the single feature regression models will be discussed first before moving on to the optimal regression model. These single feature models were constructed using direct logistic regression analysis, and include a model developed for inter-crime distances, the combined component, entry behaviours, target selection choices, property stolen, and internal behaviours.

5.5.1. The regression coefficients

Consistent with the findings presented in Table 5.1, the signs of the logit coefficients in the single feature models indicate that unlinked crime pairs tend to be characterised by larger inter-crime distances than do linked crime pairs (logit=-0.30), but have lower across-crime similarity scores for the combined component (logit=+2.83), entry behaviours (logit=+1.19), target selection choices (logit=+0.77), property stolen (logit=+0.86), and internal behaviours (logit=+0.53).

To determine what these logit coefficients actually mean, each of them was exponentiated after multiplying the coefficient by 0.10. In relation to inter-crime distances, this meant that the odds of two crimes being linked in the reduced Dorset sample would be multiplied by 0.97, which would reduce them. For the combined component, entry behaviours, target selection choices, property stolen and internal behaviours, the odds would be multiplied by 1.33, 1.13, 1.08, 1.09 and 1.05 respectively. All of these values are consistent with the fact that linked crime pairs in the reduced Dorset sample are characterised by higher levels of across-crime similarity across all the residential burglary behaviours examined.

The impact of changes in the predictor variables was also assessed by examining changes in the probability that two crimes are linked. As can be seen in Table 5.3, the general effect of increasing similarity scores in the Dorset sample is an increase in the probability that two crimes are linked. However, the rate of this change is domain specific. In this table, each row indicates an increase in across-crime similarity. The changes for inter-crime distance equate to a decrease of 0.10 km per row, from 1.00 km apart to 0.00 km apart. For all other variables the changes equate to an increase of 0.10 per row, from 0.00 to 1.00.

Table 5.3. Estimated probabilities as a function of similarity

Increases in across-crime similarity	Unit	Distance	Combined	Entry	Target	Property	Internal
	1	0.12	0.02	0.02	0.02	0.03	0.03
	2	0.12	0.03	0.02	0.03	0.03	0.03
	3	0.13	0.03	0.03	0.03	0.03	0.03
	4	0.13	0.04	0.03	0.03	0.04	0.03
	5	0.13	0.06	0.03	0.03	0.04	0.03
	6	0.14	0.08	0.04	0.04	0.04	0.04
	7	0.14	0.10	0.04	0.04	0.05	0.04
	8	0.15	0.12	0.05	0.04	0.05	0.04
	9	0.15	0.16	0.05	0.04	0.05	0.04
	10	0.15	0.20	0.06	0.05	0.06	0.04

5.5.2. Predictive accuracy and goodness-of-fit

Consistent with the findings in Table 5.1, the logistic regression models are also found to have high levels of predictive accuracy and high degrees of fit with the data. Their predictive accuracy is indicated by significant levels of Wald's statistic. The values of Wald's statistic indicate that inter-crime distances are the most significant predictors (W=71.45, df=1, $p \leq 0.001$), followed by the combined component (W=19.92, df=1, $p \leq 0.001$), entry behaviours (W=11.21, df=1, $p \leq 0.001$), target selection choices (W=5.64, df=1, $p \leq 0.01$), property stolen (W=5.47, df=1, $p \leq 0.01$), and internal behaviours (W=4.05, df=1, $p \leq 0.05$).

Model fit is indicated by the R^2 values and X^2 values. The values of R^2 indicate that the regression model including inter-crime distances explains the highest proportion of variance in the criterion variable (R^2=0.22), followed by the combined component (R^2=0.02), entry behaviours (R^2=0.01), target selection choices (R^2=0.01), property stolen (R^2=0.00), and internal behaviours (R^2=0.00). The X^2 goodness-of-fit statistics also support this. The regression model including inter-crime distances has the highest X^2 value (X^2=251.91, df=1, $p \leq 0.001$). This is followed by the combined component (X^2=18.19, df=1, $p \leq 0.001$), entry behaviours (X^2=10.65, df=1, $p \leq 0.001$), target selection choices (X^2=5.31, df=1, $p \leq 0.05$), property stolen (X^2=4.86, df=1, $p \leq 0.05$), and internal behaviours (X^2=3.79, df=1, $p \leq 0.05$).

5.6. The multiple feature model

In addition to constructing single feature regression models, an optimal regression model was constructed. This model consists of the residential burglary behaviours that combine to have the highest level of predictive power (the combined component was not included in this analysis). A summary of the optimal regression model is also presented in Table 5.2.

From this table, it can be seen that the optimal model includes 3 of the 5 predictor variables. Specifically, the model includes inter-crime distances, entry behaviours and internal behaviours, while leaving out target selection choices and property stolen. As indicated by Wald's statistic, inter-crime distances unsurprisingly have the most predictive power in this optimal model, followed by entry behaviours and internal behaviours. As indicated by the R^2 and X^2 values, the optimal model also unsurprisingly explains a higher proportion of variance in the criterion variable than any single feature model (R^2=0.24) and fits the data better (X^2=269.70, df=3, $p \leq 0.001$).

5.6.1. The redundancy of target selection choices and property stolen

Why, when target selection choices and property stolen were significant predictors in isolation were they not incorporated as predictors in the optimal model? And why, when target selection choices and property stolen were more predictive than internal behaviours, were internal behaviours included in the optimal model? The answer to these questions can be best understood by considering the inter-correlations between the predictor variables as well as the correlations and partial correlations between the predictor variables and the criterion variable.

As indicated in Table 5.4, many of the predictor variables are significantly correlated with one another[6]. As a result, it is unlikely that each predictor variable will uniquely account for a significant portion of the variance in the criterion variable, which would have enabled them all to be included in the optimal model. The correlations presented in Table 5.5 support this argument. These correlations show that while each predictor variable is significantly correlated, only three variables remain significantly correlated with the criterion variable when the effects of all other variables are removed. The variables that remain correlated after controlling for the effects of the other predictor variables include inter-crime distances, entry behaviours and internal behaviours. This helps explain why these three variables, but not the other two, form the optimal regression model.

[6]Correlations in Tables 5.4 and 5.5 have been rounded.

Table 5.4. Inter-correlations between the predictor variables

Variables	Distance	Entry	Internal	Property	Target
Distance	--	0.05**	0.04**	-0.01	-0.03*
Entry		--	0.05***	0.03*	0.12***
Internal			--	0.06***	0.02
Property				--	0.06***
Target					--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 5.5. Correlations between the predictors and the criterion variable

Predictor variables	Zero-order correlations	Partial correlations
Distance	-0.15***	-0.15***
Entry	0.05***	0.05***
Internal	0.03**	0.03*
Property	0.03**	0.03
Target	0.03**	0.02

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

5.7. ROC analysis

In the previous section, the logistic regression models indicated that a range of residential burglary behaviours could be used with a relatively high degree of accuracy when predicting whether crime pairs are unlinked or linked. However, the accuracy of the models did vary depending on the behaviours used. Of the single feature models, the one that included inter-crime distances was by far the most effective, though this model could be improved slightly by combining inter-crime distances with entry behaviours and internal behaviours.

As in the previous chapter, ROC analysis was carried out on these logistic regression models in order to obtain separate measures of consistency and discrimination, as well as some indication of model validity. To carry out this analysis, the regression models presented in Table 5.2 were used to calculate estimated probabilities for every possible crime pair in the Dorset test sample. These probabilities were then used to construct separate ROC graphs for each single feature regression model as well as the optimal regression model.

5.8. Single feature ROC graphs

The single feature ROC graphs, along with their AUCs (and *p*-values), standard errors, and 95% confidence intervals are presented in Figure 5.1. These ROC graphs correspond to the single feature regression models presented in Table 5.2 once they had been applied to each and every crime pair in the test sample.

5.8.1. The AUC as a measure of spatial and behavioural consistency

Consistent with the analysis of data in the experimental sample, each of the ROC curves in Figure 5.1 indicate that residential burglary behaviours in Dorset are expressed in a consistent fashion, though not all are consistent beyond what would be expected by chance. The ROC graphs also indicate that certain behaviours are exhibited more consistently than others are. Again, inter-crime distances are the most consistent feature in Dorset (AUC=0.89), followed by the combined component (AUC=0.67), entry behaviours (AUC=0.64), target selection choices (AUC=0.59), internal behaviours (AUC=0.57), and property stolen (AUC=0.52).

The ordering of the predictor variables based on their AUCs is very similar to the ordering based on the previous logistic regression analysis. Indeed, only internal behaviours and property stolen have switched places. Table 5.6 presents results showing which of these curves differ significantly from one another. For a more detailed description and explanation of these significance tests see Appendix B.

Figure 5.1. Single feature ROC graphs for Dorset residential burglary data

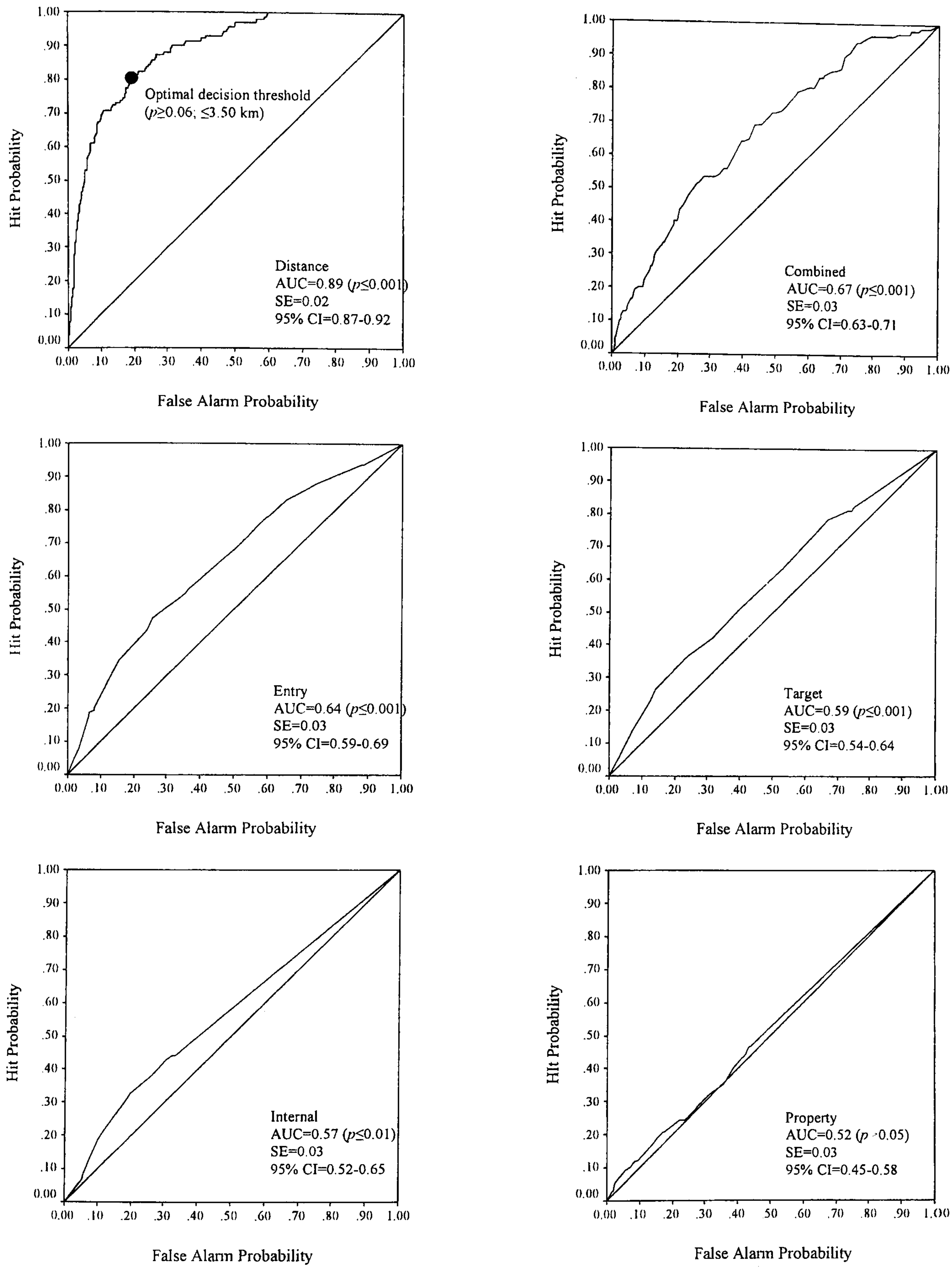


Table 5.6. Differences between the ROC curves in Figure 5.1

Variables	Distance	Combined	Entry	Target	Internal	Property
Distance	--	$p \leq 0.001$	$p \leq 0.001$	$p \leq 0.001$	$p \leq 0.001$	$p \leq 0.001$
Combined		--	*n.s.*	$p \leq 0.001$	$p \leq 0.05$	$p \leq 0.001$
Entry			--	*n.s.*	*n.s.*	$p \leq 0.001$
Target				--	*n.s.*	*n.s.*
Internal					--	*n.s.*
Property						--

5.8.2. Operating points as a measure of spatial discrimination

As was the case in the previous chapter, the ROC graphs in Figure 5.1 suggest that where the decision threshold is placed can have a serious impact on discrimination accuracy. Youden's index was calculated in order to identify an optimal threshold for each of the ROC graphs. For inter-crime distances, the optimal threshold is $p \geq 0.06$ (≤ 3.50 km). The optimal thresholds for the combined component, entry behaviours, target selection choices, internal behaviours and property stolen are $p \geq 0.03$ (≥ 0.30), $p \geq 0.03$ (≥ 0.36), $p \geq 0.03$ (≥ 0.35), $p \geq 0.03$ (≥ 0.24), and $p \geq 0.03$ (≥ 0.18) respectively. As explained in the previous chapter, one of the reasons the optimal decision thresholds are all so low (with respect to the p-values) is because of the large discrepancy between the number of unlinked and linked crime pairs in the Dorset sample.

5.8.3. Measuring improvements in discrimination accuracy

When a variety of burglary behaviours are used to generate multiple ROC curves, as in the present case, it is relatively easy to determine which behaviours should be used for discrimination purposes. As long as the definitions proposed in Chapter 1 are adopted, the behaviours expressed most consistently are the best candidates, since a higher number of correct linking decisions will be made using these behaviours if the number of incorrect decisions is held constant. However, it would also be useful to know exactly how much better one behavioural domain is over another in more precise terms.

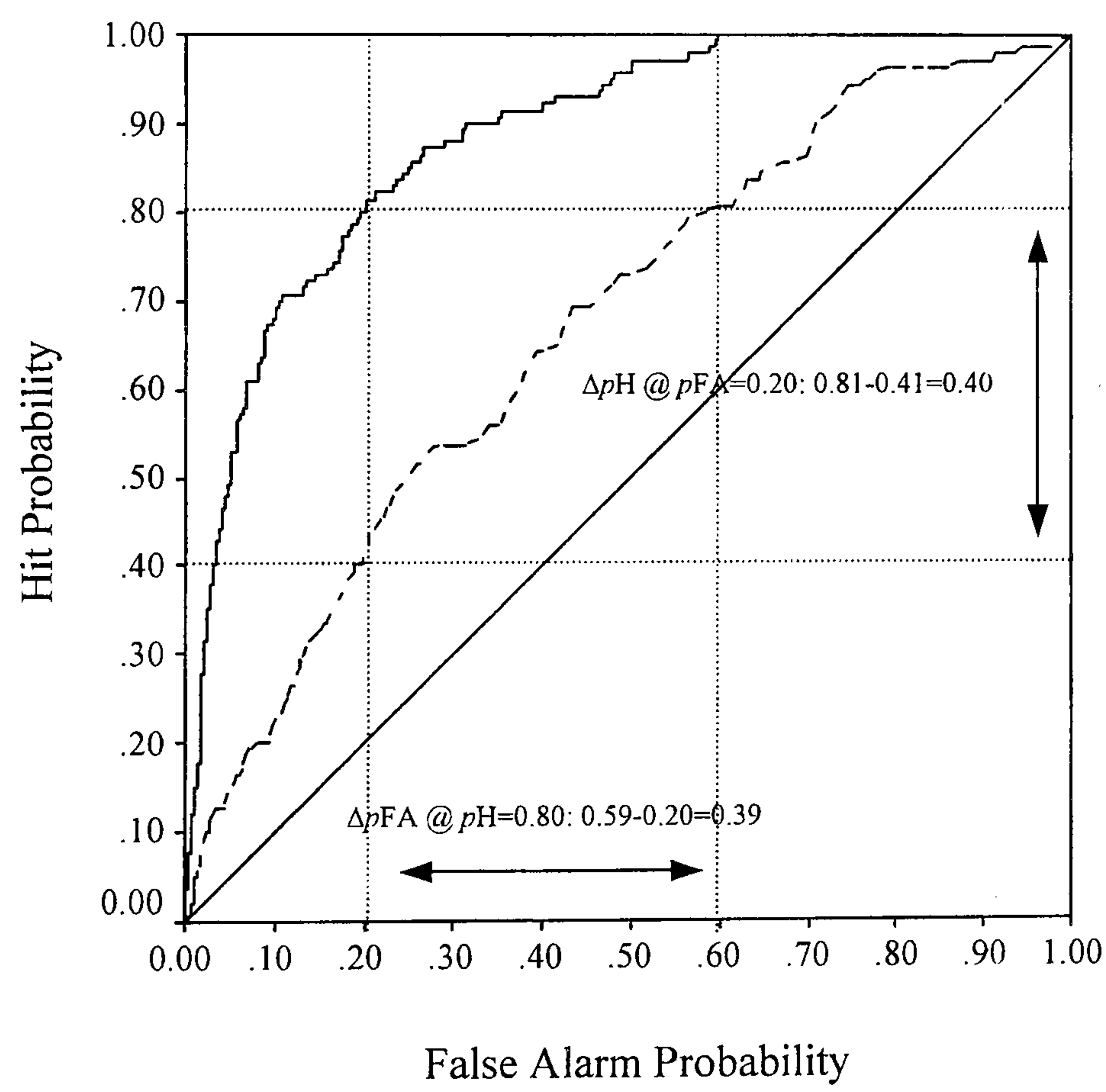
Improvements in discrimination accuracy can come in one of two ways, either as an increase in the number of hits that result from using one behavioural domain over another, or as a decrease in the number of false alarms. A measure of improvement, then, can be derived from an examination of how many more hits are achieved at a pre-specified rate of false alarms when using different burglary behaviours, or by examining how many less false alarms are achieved at a pre-specified rate of hits (Swets *et al.*, 2000a).

To illustrate this point, consider the top two ROC graphs in Figure 5.1, which have been combined to form Figure 5.2. The two ROC curves in Figure 5.2 correspond to inter-crime distances (the solid ROC curve on top) and the combined component (the dashed ROC curve on bottom). The height of these curves suggests that both features are exhibited in a relatively consistent fashion and therefore both will be effective at discriminating between unlinked and linked crime pairs. Inter-crime distances, however, are clearly more effective in general. The question is, how much better?

To answer this questions, consider a police force who sets a limit on the rate of false alarms they can make during burglary investigations, say *p*FA=0.20. At a rate of *p*FA=0.20 in Figure 5.2, it is possible to get a hit rate of 0.41 when using the combined component (as indicated by the bottom horizontal line in Figure 5.2). However, at the exact same *p*FA, it is possible to get a hit rate of 0.81 when using inter-crime distances (as indicated by the top horizontal line in Figure 5.2). Thus, at a pre-specified rate of *p*FA equal to 0.20, an additional 40 hits (0.81-0.41=0.40) can be made for every 100 crime pairs encountered if inter-crime distances are used as the basis for making linking decisions instead of the combined component.

Alternatively, a police force may demand a high rate of hits in burglary investigations, say *p*H=0.80. At a rate of *p*H=0.80 in Figure 5.2, it is possible to get a false alarm rate of 0.59 when using the combined component (as indicated by the right vertical line in Figure 5.2). However, at the same *p*H, it is possible to get a false alarm rate of 0.20 when using inter-crime distances (as indicated by the left vertical line in Figure 5.2). Thus, at a pre-specified rate of *p*H equal to 0.80, 39 false alarms (0.59-0.20=0.39) could be avoided for every 100 crime pairs encountered if inter-crime distances were used instead of the combined component.

Figure 5.2. Improvements in discrimination accuracy
(solid ROC curve: distance, dashed ROC curve: combined)

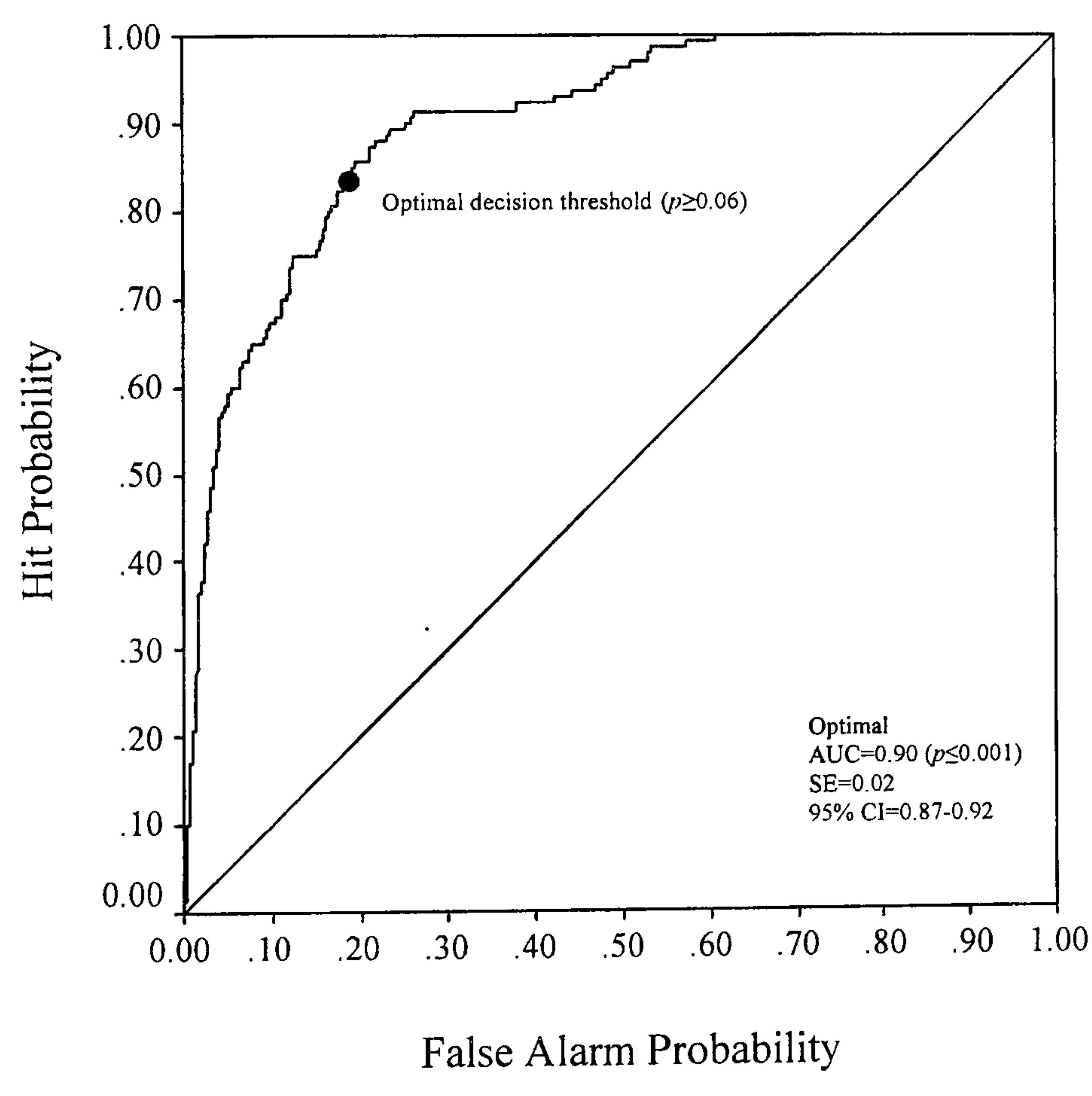


5.9. The multiple feature ROC graph

In order to get some indication of how valid the optimal regression model is it was also used to construct a ROC graph using data from the test sample. This ROC graph, along with its AUC (and *p*-value), standard error, and 95%

confidence interval, is presented in Figure 5.3. As with the ROC graphs presented in Figure 5.1, this ROC graph supports the previous regression analysis. Not only does the optimal regression model result in a ROC curve that is significantly more accurate than chance, it also results in a ROC curve that is slightly more accurate than any of the single feature ROC curves. Specifically, the AUC for the ROC graph depicted in Figure 5.3 is 0.90. This is not significantly greater than the AUC found for inter-crime distances, but it is significantly greater than all the other AUCs at a level of $p \leq 0.001$. The optimal decision threshold for this multiple feature ROC curve, as determined using Youden's index, is equal to $p \geq 0.06$.

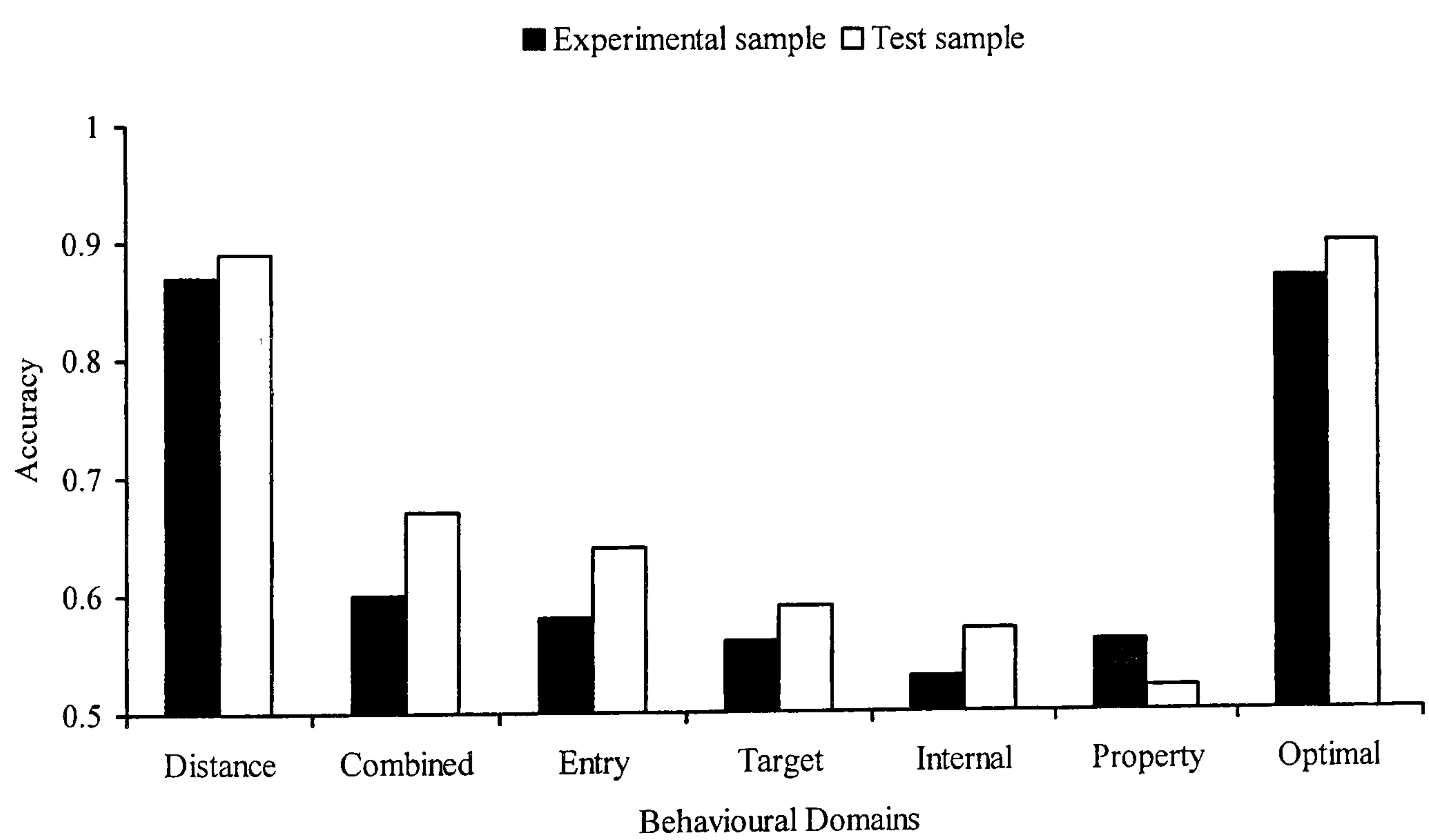
Figure 5.3. Multiple feature ROC graph for Dorset residential burglary data



5.10. Validating the empirical ROC curves

When the logistic regression models developed using data from the experimental sample are applied to data from the test sample, the ROC curves that result from the most accurate regression models have AUCs that are significantly greater than chance. This suggests that these logistic regression models do generalise to crimes beyond those used to construct the original models. In addition, the AUCs associated with each ROC curve generally correspond to how accurate the regression models are. Only internal behaviours and property stolen switch places. However, the validity of these models can be tested more directly by constructing ROC curves using data from the experimental sample and comparing these to the ROC curves in Figure 5.1. This is done in Figure 5.4 where it can be seen that the logistic regression models perform slightly better using data from the test sample. Only in relation to property stolen does the AUC in the experimental sample exceed the AUC obtained from the test sample.

Figure 5.4. Comparison of AUCs across the experimental and test samples



5.10.1. External discrimination trials

In addition to this form of validation, a number of external discrimination trials were also carried out. Again, the goal was to determine whether the values of *p*H and *p*FA associated with optimal decision thresholds in Figure 5.1 correspond to the values of *p*H and *p*FA obtained across discrimination trials. In this case, only the ROC curve associated with inter-crime distances was tested since this was the most effective discriminator for Dorset burglaries. Recall the values of pH and pFA achieved at the optimal decision threshold of $\leq$3.50 km. The rates are 0.82 and 0.22 respectively. If the ROC graph for inter-crime distances provides a valid representation of how accurately discriminations can be made, it should be possible to achieve similar levels of accuracy across random trials when this decision threshold is used.

Using the same method as in the previous chapter, the results of these trials are presented in Table 5.7. Across 5 randomly selected samples of 1000 crime pairs, the average hit and false alarm rates were 0.79 and 0.19 respectively. Across 5 randomly selected samples of 10000 crime pairs, the average hit and false alarm rates were 0.70 and 0.18 respectively. These rates appear to generally correspond to the predicted values, though the hit rates generated from the larger samples are not quite so close. In addition, on every trial there was a highly significant association between predictions and reality, as indicated by the X^2 values presented in the last column of the table.

5.11. Chapter summary

In this chapter, the behaviours of residential burglars in the county of Dorset were explored. This extended the analysis of residential burglary behaviour presented in the previous chapter in two important ways. First, the area where these burglaries were collected from is much more rural than London is. Second, a variety of burglary behaviours could be examined using the Dorset data, beyond the inter-crime distances examined from London. These behavioural domains included entry behaviours, target selection choices, internal behaviours and property stolen. Descriptive statistics provided support for the hypothesis that the behaviours making up these behavioural domains are exhibited in a consistent fashion by burglars in Dorset. Highly significant differences were found between the similarity scores for unlinked and linked crime pairs for every behavioural domain, and all of these differences were in the expected direction.

Logistic regression analysis was run on the similarity scores associated with each behavioural domain as a way of calculating predicted probabilities that crime pairs were linked. These probabilities, in turn, formed the basis of ROC analysis.

Consistent with the descriptive analysis, ROC analysis indicated that inter-crime distances are the most consistent aspect of burglary behaviour in Dorset, with much lower levels of consistency found for the other domains. In line with this, high levels of discrimination accuracy could also be achieved when using inter-crime distances. Using a distance threshold of ≤3.50 km, 82% of linked burglaries were correctly classified as linked and only 22% of unlinked burglaries were incorrectly classified as linked. Similar results emerged across discrimination trials, suggesting that the ROC procedure has an adequate level of predictive accuracy in Dorset.

As was the case in the previous chapter, these results confirm that residential burglars are spatially limited in terms of the distances that exist between their crime site locations. This apparently is the case even when the crimes are committed in a largely rural area where burglars would be expected to travel further distances (Van Koppen & Jansen, 1998). The results presented in this chapter also suggest that the level of discrimination accuracy that can be achieved using crimes scene behaviors is often remarkably low compared to the accuracy associated with inter-crime distances. Such a finding sits in contrast to claims made by Green and his colleagues (1976), particularly their suggestion that crime scene behaviours can form a reliable basis for distinguishing between burglaries committed by different offenders. This discrepancy may indicate that the findings reported in that early study are specific to the very small number of burglars that were examined.

Table 5.7. Validation trials for Dorset residential burglary data

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.06$ (≤ 3.50 km)	1000	0.89 (24)	0.11 (3)	0.81 (791)	0.19 (182)	79.12 (1)***
	$p \geq 0.06$ (≤ 3.50 km)	10000	0.70 (232)	0.30 (101)	0.82 (7934)	0.18 (1733)	545.87 (1)***
2	$p \geq 0.06$ (≤ 3.50 km)	1000	0.80 (19)	0.20 (5)	0.80 (783)	0.20 (193)	49.46 (1)***
	$p \geq 0.06$ (≤ 3.50 km)	10000	0.69 (216)	0.31 (98)	0.82 (7950)	0.18 (1736)	500.93 (1)***
3	$p \geq 0.06$ (≤ 3.50 km)	1000	0.67 (18)	0.33 (9)	0.81 (792)	0.19 (181)	38.08 (1)***
	$p \geq 0.06$ (≤ 3.50 km)	10000	0.68 (220)	0.32 (104)	0.82 (7979)	0.18 (1679)	513.18 (1)***
4	$p \geq 0.06$ (≤ 3.50 km)	1000	0.77 (20)	0.23 (6)	0.81 (791)	0.19 (183)	52.90 (1)***
	$p \geq 0.06$ (≤ 3.50 km)	10000	0.69 (215)	0.31 (97)	0.82 (7973)	0.18 (1715)	508.90 (1)***
5	$p \geq 0.06$ (≤ 3.50 km)	1000	0.80 (20)	0.20 (5)	0.82 (795)	0.18 (180)	57.69 (1)***
	$p \geq 0.06$ (≤ 3.50 km)	10000	0.71 (226)	0.29 (91)	0.82 (7946)	0.18 (1737)	553.86 (1)***
Average	$p \geq 0.06$ (≤ 3.50 km)	1000	0.79 (20.20)	0.21 (5.60)	0.81 (790.40)	0.19 (183.80)	--
	$p \geq 0.06$ (≤ 3.50 km)	10000	0.70 (221.80)	0.30 (98.20)	0.82 (7956.40)	0.18 (1723.60)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

CHAPTER 6
THE BEHAVIOUR OF SERIAL BURGLARS IN OLDHAM

6.1. Introduction

In contrast to previous chapters, where burglary behaviour was examined at a force-wide level, this chapter explores the behaviour of burglars committing crimes in Oldham, which is one division of the Greater Manchester Police Service. Furthermore, where the focus was specifically on residential burglary in previous chapters, this chapter extends the examination to include commercial burglary as well. As was the case with Dorset burglaries, the data collected from Oldham allows a variety of different aspects of burglary behaviour to be examined. These include the spatial aspect of residential and commercial burglary, as well as entry behaviours, target selection choices and property stolen. The objectives in this chapter are to determine whether these behaviours are expressed in a consistent fashion by residential and commercial burglars in the Oldham division, and if so, whether an analysis of these behaviours can form a reliable basis for distinguishing between crimes committed by different offenders.

Unlike residential burglary, commercial burglary has rarely been examined. Although there are some notable exceptions (e.g., Walsh, 1986), the fact that residential burglaries are so potentially traumatic for the victim has meant that a disproportionate amount of time has been spent trying to understand this crime (e.g., Bennett & Wright, 1984; Cromwell, Olson & Avary, 1991; Maguire, 1982; Walsh, 1980; Wright & Decker, 1994). Due to this lack of research, it is difficult to predict whether the patterns of consistency and discrimination found in previous chapters will also be found in relation to commercial burglary. On the one hand, some research suggests that many of the behavioural patterns exhibited by commercial burglars are similar to those exhibited by residential burglars (e.g., Walsh, 1986). On the other hand, some research draws attention to the many important differences that exist between residential and commercial burglars, especially in relation to their spatial and temporal behaviour (e.g., Butler, 1994; Capone & Nichols, 1976; Van Koppen & Jansen, 1998; Wiersma, 1996).

6.1.1. The area

The residential and commercial burglaries included in the Oldham sample were committed between January 1999 and December 1999 in the Oldham division of the Greater Manchester Police Service. This division is located in the northern part of Greater Manchester and borders on Lancashire and West Yorkshire. The division covers an area of approximately 141 km^2, has a population of about 220000 persons, and a population density of approximately 1551 persons/km^2.

The Oldham division consists of 420 police officers. These officers patrol the 2 sub-divisions making up the Oldham division, which include Chatterton and Oldham. In turn, these 2 sub-divisions include a number of different geographic areas, including Chadderton, Failsworth, Oldham, Royton, Shaw and Uppermill. These areas differ in terms of their demographics and topography. As a result it is difficult to provide a general description of the Oldham division. For example, the western part of the division is densely population and heavily urbanised, whereas the eastern part largely consists of rural and tourist areas.

At the divisional level in 2001, the frequency of reported residential and commercial burglary was 3247 and 2829 respectively, with corresponding clear-up rates of 10% and 9%. However, the incidents of burglary vary across the division and are traditionally concentrated in the southwest, where very high rates of social and economic disadvantage are also experienced. In terms of the frequency of reported crimes, residential and commercial burglaries are not the most serious problem in the Oldham division. Compared with burglary, for example, rates for criminal damage and crimes against vehicles are much higher. However, the high volume of burglary, along with the relatively low clear-up rates, suggests that residential and commercial burglaries represent a serious problem for the police in Oldham division.

6.1.2. The data

As was the case with the data collected from London and Dorset, the information pertaining to spatial behaviour in both the residential and commercial burglary sample consists of geo-coded x and y coordinates indicating the position of each burglary location to the nearest metre. For all other behaviours, the information was coded in dichotomous form, with a 1 indicating that a particular behaviour was present and a 0 indicating it was absent. A more detailed list of the behaviours included in the residential and commercial burglary samples can be found in Appendix F and G.

The entire residential burglary sample collected from Oldham consists of 36 serial burglars responsible for 150 crimes. The offence series range in length from 2 crimes to 22 crimes. For the purpose of the analysis presented in this chapter, however, a smaller subset of residential burglaries was selected from this sample. This smaller sample will henceforth be referred to as the reduced Oldham residential burglary sample. Specifically, 2 crimes from each offender were randomly selected from the entire sample to form a reduced Oldham residential burglary sample of 72 crimes.

The entire commercial burglary sample collected from Oldham consists of 43 serial burglars responsible for 135 crimes. The offence series range in length from 2 crimes to 9 crimes. For the purpose of the analysis presented in this chapter, a smaller subset of commercial burglaries was also selected and will henceforth be referred to as the reduced Oldham commercial burglary sample. As with the residential burglary sample, 2 crimes from each offender were randomly selected from the entire sample to form a reduced Oldham commercial burglary sample of 86 crimes.

As in the previous chapters, this selection procedure was carried out to control for any bias that may be introduced into the analysis as a result of sampling from an uneven distribution of offences across offenders. To ensure the results do in fact generalise to the larger samples, random, large-scale discrimination trials were once again carried out and are presented at the end of the chapter.

6.2. Calculating spatial and behavioural similarity scores

In order to examine the behaviour of residential and commercial burglars committing crimes in Oldham, spatial and behavioural similarity scores were calculated. As before, the spatial similarity scores for both samples consist of inter-crime distances, calculated by entering the x and y coordinates from each of the samples into *S-LINK*. The behavioural similarity scores once again consist of Jaccard coefficients and were calculated for each of the behavioural domains by entering the dichotomous data from both samples into *B-LINK*. All of the generated crime pairs from *S-LINK* and *B-LINK* were then defined as unlinked or linked, based on who was known to have committed the crimes. This procedure resulted in 2520 unlinked crime pairs and 36 linked crime pairs for the reduced residential sample, and 3614 unlinked crime pairs and 41 linked crime pairs for the reduced commercial sample.

6.3. A descriptive analysis of the spatial and behavioural similarity scores

As in previous chapters, the first step in examining issues of consistency and discrimination was to calculate descriptive statistics. These were calculated across all unlinked and linked crime pairs. As before, it was expected that similarity scores calculated across unlinked crime pairs would be low relative to similarity scores calculated across linked crime pairs. If this were found to be the case, it would provide support that the behaviours exhibited by burglars in Oldham are expressed consistently over time, thus making it possible to discriminate between crimes committed by different offenders.

6.3.1. A descriptive analysis of residential burglary behaviours

The descriptive analysis for residential burglary is presented in Table 6.1. This table includes the mean values of similarity scores, along with their ranges and standard deviations. In addition, the results from t-tests are provided. As indicated by the mean values, unlinked crime pairs in residential burglary consistently have lower similarity scores compared to linked crime pairs. However, cases exist where high levels of similarity are found for unlinked crime pairs as well as low levels of similarity for linked crime pairs. In fact, there are a number of cases where unlinked crime pairs have a higher maximum similarity score than linked crime pairs. As a result, and in contrast to previous chapters, significant differences are not always found between the similarity scores. The t-tests presented in Table 6.1 indicate that a significant difference exists in 3 cases out of the 5. A highly significant difference is found for inter-crime distances ($p \leq 0.001$), whereas significant differences are found for the combined component and property stolen at a level of $p \leq 0.05$.

Table 6.1. Summary of the Oldham residential burglary data

	Unlinked crime pairs (n=2520 pairs)			Linked crime pairs (n=36 pairs)			
Variables	M	Range	SD	M	Range	SD	t
Distance	3.16	0-8.48	1.70	1.16	0-4.48	1.22	7.01***
Combined	0.24	0-1.00	0.15	0.30	0-0.54	0.13	2.31*
Property	0.15	0-1.00	0.21	0.23	0-1.00	0.24	2.30*
Target	0.43	0-1.00	0.32	0.51	0-1.00	0.31	1.42
Entry	0.19	0-1.00	0.22	0.21	0-0.67	0.20	0.61

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

6.3.2. A descriptive analysis of commercial burglary behaviours

The descriptive analysis for commercial burglary is presented in Table 6.2. As in Table 6.1, this table includes the mean values of similarity scores, along with their ranges and standard deviations. In addition, the results from t-tests are provided. As indicated by the mean values, unlinked crime pairs in commercial burglary again have consistently lower similarity scores compared to linked crime pairs. Furthermore, as in all previous analyses, cases exist where high levels of similarity are found for unlinked crime pairs, as well as low levels of similarity for linked crime pairs. Unlike the analysis of residential burglary behaviours in Oldham, however, significant differences are found in every case. The t-tests presented in Table 6.2 indicate that all differences are highly significant at the level of $p \leq 0.001$.

Table 6.2. Summary of the Oldham commercial burglary data

	Unlinked crimes pairs (n=3614 pairs)			Linked crime pairs (n=41 pairs)			
Variables	M	Range	SD	M	Range	SD	t
Distance	4.13	0-12.47	2.53	1.37	0-4.39	1.33	6.98***
Combined	0.27	0-1.00	0.12	0.35	0-0.83	0.20	6.60***
Target	0.34	0-1.00	0.18	0.51	0-1.00	0.34	5.67***
Entry	0.17	0-1.00	0.21	0.31	0-1.00	0.29	4.20***
Property	0.19	0-1.00	0.24	0.32	0-1.00	0.30	3.50***

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

6.3.3. Differences between residential and commercial burglary

Clearly, unlinked and linked crime pairs can be differentiated to different degrees in cases of residential and commercial burglary. Whereas only 3 of the 5 comparisons differed significantly in the analysis of residential burglaries, all 5 comparisons differed significantly in the analysis of commercial burglaries. To determine if there were any other potentially important differences, a series of 2-way ANOVAs were carried out. The two independent variables in these analyses were the status of the crime pair (unlinked versus linked) and the type of crime (residential versus commercial). The dependent variable was the across-crime similarity scores. The results of this analysis are presented in Table 6.3.

Table 6.3. ANOVAs in relation to status and type of burglary

Variable	Status	Type	Interaction
Distance	F(1,6207)=87.02***	F(1,6207)=5.37*	F(1,6207)=2.29
Combined	F(1,6207)=36.25***	F(1,6207)=0.42	F(1,6207)=5.36*
Target	F(1,6207)=17.47***	F(1,6207)=2.32	F(1,6207)=2.28
Property	F(1,6207)=16.50***	F(1,6207)=6.07**	F(1,6207)=0.91
Entry	F(1,6207)=10.95***	F(1,6207)=3.08	F(1,6207)=5.87*

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

The results presented in Table 6.3 support the findings presented in Tables 6.1 and 6.2 in as much as a main effect is found for the status factor. In every case, the average similarity scores are lower for unlinked crime pairs. In addition, for two aspects of burglary behaviour, inter-crime distances and property stolen, a main effect is also found for the type of crime factor. In both of these cases, the average similarity scores are higher in the case of commercial burglary. Thus, Oldham commercial burglars as a group appear to be more geographically focused than residential burglars and they steal items that are more similar. Also interesting is the fact that the gap in similarity scores between unlinked and linked crime pairs is consistently greater in the case of commercial burglary, though in only two cases does this lead to a significant interaction effect.

6.4. Logistic regression analysis

The descriptive analysis presented in the previous section provides support that burglars committing crimes in Oldham division express a range of offence behaviours in a consistent fashion. This, in turn, suggests it may be possible to use these behaviours to discriminate between crimes committed by different offenders. However, as in previous chapters, the fact that there is a degree of overlap between unlinked and linked similarity scores suggests that the level of consistency exhibited by burglars in Oldham will not be absolute, and that discrimination accuracy will not be perfect. Furthermore, the fact that the degree of overlap varies depending on the behavioural domain and type of crime considered suggests that consistency and discrimination will be both domain and crime specific.

As done in the previous chapters, to examine these issues more directly, single feature regression models were developed from the residential and commercial burglary data, as well as optimal logistic regression models. As was previously discussed, the development of these models takes the analysis of burglary behaviour beyond a simple examination of average across-crime similarity scores. These models provide further evidence of the degree to which residential and commercial burglaries in Oldham can be discriminated from one another, and they provide the necessary data for ROC analysis.

6.4.1. Validation datasets

The datasets containing all the crime pairs generated from the reduced Oldham samples were split in half for the purpose of validating the logistic regression models. The logistic regression models were developed on the experimental samples and tested for generalisation on the test samples. Tables 6.4 and 6.5 contain a summary of the logistic regression models. A range of information is provided in these tables, including the model coefficients and standard errors (constant and logit), an indicator of predictive accuracy for each predictor variable (Wald's statistic), and indices of general model fit (R^2 and X^2).

6.5. Logistic regression models for residential burglary

Each of the single feature regression models constructed from the residential burglary sample will be discussed first before moving on to the optimal regression model. These single feature models were constructed using direct logistic regression analysis, and include a model developed for inter-crime distances, the combined component, entry behaviours, target selection choices, and property stolen.

Table 6.4. Logistic regression models for Oldham residential burglary [a,b]

Model[c]	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-1.64 (0.38)	-1.49 (0.31)	23.89 (1)***	0.24	43.70 (1)***
Combined	-4.75 (0.49)	1.89 (1.50)	1.60 (1)	0.01	1.51 (1)
Property	-4.40 (0.30)	0.84 (0.93)	0.82 (1)	0.00	0.74 (1)
Entry	-4.43 (0.34)	0.87 (1.05)	0.68 (1)	0.00	0.65 (1)
Target	-4.48 (0.42)	0.51 (0.71)	0.50 (1)	0.00	0.49 (1)
Optimal	-1.64 (0.38)	--	--	0.24	43.70 (1)***
Distance	--	-1.49 (0.31)	23.89 (1)***	--	--

[a] Criterion variable (unlinked crime pair=0, linked crime pair=1); [b] Sample size=1278 crime pairs (1260 unlinked, 18 linked); [c] Direct logistic regression analysis was used to construct the single feature models; forward stepwise regression analysis was used to construct the optimal model (inclusion criteria: $p \leq 0.05$); *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

6.5.1. The regression coefficients

Consistent with the findings presented in Table 6.1, the signs of the logit coefficients in the single feature models indicate that unlinked crime pairs in residential burglary tend to be characterised by larger inter-crime distances compared to linked crime pairs (logit=-1.49), but have lower across-crime similarity scores for the combined component (logit=+1.89), property stolen (logit=+0.84), entry behaviours (logit=+0.87) and target selection choices (logit=+0.51).

As before, the effect of increasing the similarity scores by 0.10 units was examined by multiplying each logit coefficient by 0.10 and exponentiating it. In relation to inter-crime distances, this meant that the odds of two crimes being linked would be multiplied by 0.86, which would reduce them. For the combined component, property stolen, entry behaviours and target selection choices, the odds would be multiplied by 1.21, 1.09, 1.09 and 1.05 respectively. All these calculations are consistent with the fact that unlinked crime pairs in the reduced residential burglary sample are characterised by lower levels of across-crime similarity than linked crime pairs.

The impact of changes in the predictor variables was also assessed by examining changes in the probability that two crimes are linked. Unsurprisingly, considering the odds calculations just carried out, the general effect of increasing across-crime similarity scores is an increase in the probability that two crimes are linked. However, the rate of increase is dependent on the behavioural domain examined. For example, as the distance between two residential burglaries decreases from 1.00 km to 0.10 km in Oldham, the probability of the two crimes being linked increases from 0.04 to 0.14. However, when the similarity score for target selection choices increases from 0.10 to 1.00, the probability only increases from 0.01 to 0.02.

6.5.2. Predictive accuracy and goodness-of-fit

Also consistent with the findings presented in Table 6.1, only one of the logistic regression models developed using data from the reduced residential burglary sample was found to have a high level of predictive accuracy or a high degree of fit with the data. The values of Wald's statistic indicate that inter-crime distances have a high level of predictive power (W=23.89, df=1, $p \leq 0.001$), but all the other behavioural domains have non-significant values. The values of R^2 indicate that the regression model including inter-crime distances is the only model that explains a substantial proportion of the variance in the criterion variable (R^2=0.24). In support of this, the only model with a highly significant X^2 value is also the model including inter-crime distances (X^2=43.70, df=1, $p \leq 0.001$).

6.5.3. The multiple feature model

In addition to constructing single feature logistic regression models from the residential burglary sample, an optimal regression model was constructed (the combined component was again not included in this analysis). A summary of the optimal logistic regression model is also presented in Table 6.4. Unsurprisingly, the optimal model only includes inter-crime distances. As a result, this model has the same Wald's statistic, R^2 value and X^2 value as the single feature regression model that includes inter-crime distances.

6.6. Logistic regression models for commercial burglary

Each of the single feature regression models constructed from the reduced commercial burglary sample will be discussed first before moving on to the optimal regression model. These single feature models were again constructed using direct logistic regression analysis, and include a model developed for inter-crime distances, the combined component, entry behaviours, target selection choices, and property stolen.

Table 6.5. Logistic regression models for Oldham commercial burglary [a,b]

Model[c]	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-2.33 (0.34)	-0.90 (0.19)	23.02 (1)***	0.18	39.42 (1)***
Combined	-5.82 (0.43)	5.07 (1.10)	21.15 (1)***	0.07	15.99 (1)***
Target	-5.76 (0.45)	3.10 (0.74)	17.58 (1)***	0.06	13.85 (1)***
Property	-4.89 (0.31)	1.69 (0.67)	6.37 (1)**	0.03	5.40 (1)*
Entry	-4.84 (0.31)	1.69 (0.78)	4.70 (1)*	0.02	4.02 (1)*
Optimal	-3.35 (0.58)	--	--	0.21	44.70 (2)***
Distance	--	-0.79 (0.19)	17.63 (1)***	--	--
Target	--	1.90 (0.79)	5.87 (1)*	--	--

[a] Criterion variable (unlinked crime pair=0, linked crime pair=1); [b] Sample size=1828 (1807 unlinked, 21 linked); [c] Direct logistic regression analysis was used to construct the single feature models; forward stepwise regression analysis was used to construct the optimal model (inclusion criteria: $p \leq 0.05$); *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

6.6.1. The regression coefficients

Consistent with the findings presented in Table 6.2, the signs of the logit coefficients in the single feature models indicate that unlinked crime pairs in commercial burglary tend to be characterised by larger inter-crime distances compared to linked crime pairs (logit=-0.90), but have lower across-crime similarity scores for the combined component (logit=+5.07), target selection choices (logit= +3.10), property stolen (logit=+1.69), and entry behaviours (logit=+1.69).

To determine what these logit coefficients mean, each of them was again multiplied by 0.10 and exponentiated. In relation to inter-crime distances, it was found that for every decrease of 0.10 the odds of two commercial burglaries being linked would be multiplied by 0.91, which would reduce them. For the combined component, target selection choices, property stolen, and entry behaviours, it was found that the odds would be multiplied by 1.66, 1.36, 1.18 and 1.18 respectively. All these calculations are consistent with the fact that unlinked crime pairs in Oldham are characterised by lower levels of across-crime similarity than linked crime pairs. The probability of two crimes being linked as similarity scores increase was also examined. As before, the general effect of increasing the degree of similarity was domain-specific increases in the probability that two crimes are linked.

6.6.2. Predictive accuracy and goodness-of-fit

Also consistent with the findings presented in Table 6.2, the logistic regression models developed using data from the commercial burglary sample were all found to have high levels of predictive accuracy and high degrees of fit. The values of Wald's statistic indicate that inter-crime distances are the most significant predictors (W=23.02, df=1, $p \leq 0.001$), followed by the combined component (W=21.15, df=1, $p \leq 0.001$), target selection choices (W=17.58, df=1, $p \leq 0.001$), property stolen (W=6.37, df=1, $p \leq 0.01$), and entry behaviours (W=4.70, df=1, $p \leq 0.05$).

The values of R^2 indicate that the regression model including inter-crime distances explains the highest proportion of variance in the criterion variable

(R^2=0.18), followed by the combined component (R^2=0.07), target selection choices (R^2=0.06), property stolen (R^2=0.03), and entry behaviours (R^2=0.02). The values of X^2 indicate that the regression model including inter-crime distances has the highest degree of fit with the data (X^2=39.42, df=1, $p \leq 0.001$), followed by the combined component (X^2=15.99, df=1, $p \leq 0.001$), target selection choices (X^2=13.85, df=1, $p \leq 0.001$), property stolen (X^2=5.40, df=1, $p \leq 0.05$), and entry behaviours (X^2=4.02, df=1, $p \leq 0.05$).

6.6.3. The multiple feature model

In addition to constructing single feature regression models, an optimal regression model was constructed (the combined component was again not included in this analysis). A summary of the optimal logistic regression model is also presented in Table 6.5. In contrast to the optimal model developed from residential burglary behaviours, the optimal model developed from commercial burglary behaviours contains 2 of the 5 predictor variables. Specifically, the optimal model includes inter-crime distances and target selection choices while leaving out property stolen and entry behaviours. As indicated by Wald's statistic, inter-crime distances unsurprisingly have the most predictive power in the optimal model followed by target selection choices. As indicated by the R^2 and X^2 values, the optimal model also unsurprisingly explains a higher proportion of the variance in the criterion variable than any single feature model (R^2=0.21) and fits the data better (X^2=44.70, df=2, $p \leq 0.001$).

6.6.4. The redundancy of property stolen and entry behaviours

To examine why property stolen and entry behaviours were not incorporated as predictors in the optimal model, even though they were significant predictors in isolation, the inter-correlations between the predictor variables were examined as well as the correlations and partial correlations between the predictor variables and the criterion variable. The correlations in Table 6.6 indicate that many of the predictor variables have significant correlations with one another[7]. Thus, it is unlikely that each variable will uniquely account for a significant proportion of the variance in the criterion variable, which would have enabled them all to be

[7]The correlations in Tables 6.6 and 6.7 are rounded.

included in the optimal regression model. The correlations in Table 6.7 support this, showing that when the effects of all other variables are controlled for, the only predictor variables that remain significantly correlated with the criterion variable are inter-crime distances and target selection choices.

Table 6.6. Inter-correlations between the predictor variables

Variable	Distance	Target	Property	Entry
Distance	--	-0.15***	-0.12***	-0.04
Target		--	0.13***	0.13***
Property			--	0.17***
Entry				--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 6.7. Correlations between the predictors and the criterion variable

Predictor variables	Zero-order correlations	Partial correlations
Distance	-0.12***	-0.10***
Target	0.11***	0.08***
Property	0.06**	0.03
Entry	0.05*	0.03

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

6.7. ROC analysis

In the previous section, logistic regression models constructed from data in the experimental samples indicated that a number of residential and commercial burglary behaviours in Oldham could be used to predict whether crime pairs were unlinked or linked. However, the accuracy of these models did vary depending on the behavioural domain and the type of crime considered. In the case of residential burglary, only inter-crime distances had predictive power. In the case of commercial burglary, all behaviours had some predictive power, though inter-crime distances had the most.

As in previous chapters, ROC analysis was carried out on these logistic regression models in order obtain separate measures of consistency and discrimination for residential and commercial burglary behaviours, as well as some indication of model validity. To carry out this analysis, the regression models presented in Tables 6.4 and 6.5 were used to calculate estimated probabilities for every possible crime pair in the test samples. These probabilities were then used to construct separate ROC graphs for residential and commercial burglary behaviours, including the optimal models.

6.8. Single feature ROC graphs for residential burglary

The single feature ROC graphs for residential burglary, along with their AUCs (and *p*-values), standard errors, and 95% confidence intervals, are presented as the first 5 graphs in Figure 6.1. These ROC graphs correspond to the single feature regression models presented in Table 6.4 once they had been applied to each and every crime pair in the residential test sample.

6.8.1. The AUC as a measure of spatial and behavioural consistency

Consistent with the analysis of data in the experimental sample, each of the ROC curves in Figure 6.1 indicate that residential burglary behaviours in Oldham are expressed in a consistent fashion, though not all are consistent beyond what would be expected by chance. In addition, the ROC graphs confirm that certain behaviours are exhibited more consistently than others are. According to the ROC graphs, inter-crime distances are the most consistent feature (AUC=0.80), followed by the combined component (AUC=0.65), property stolen (AUC=0.64), target selection choices (AUC=0.60), and entry behaviours (AUC=0.53). Thus, the ordering of the predictor variables based on their AUCs is very similar to the ordering based on the previous logistic regression analysis, although target selection choices and entry behaviours have switched places. Table 6.8 presents results showing which of these curves differ significantly from one another.

Figure 6.1. ROC graphs for Oldham residential burglary data

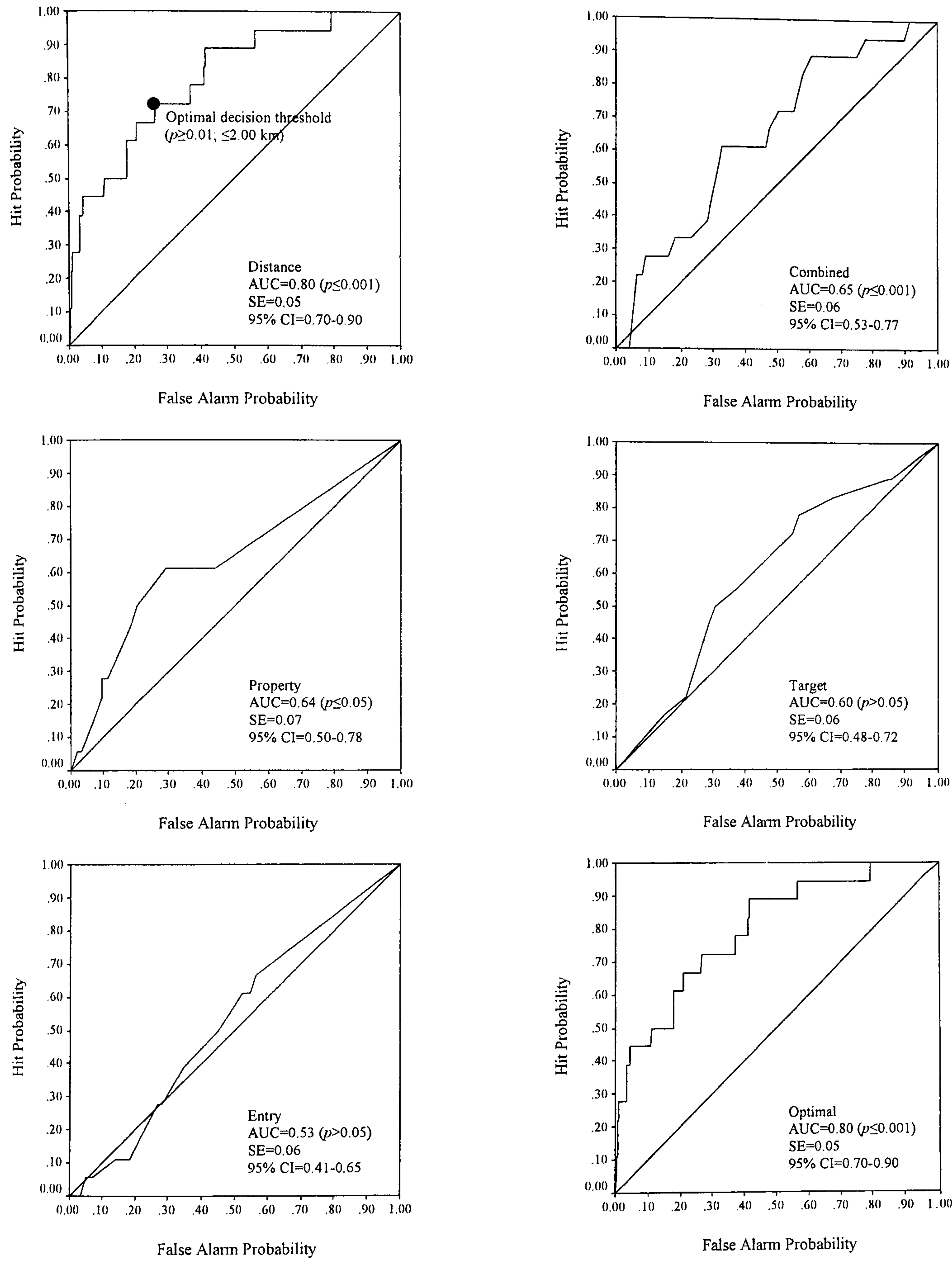


Table 6.8. Differences between the ROC curves in Figure 6.1

Variables	Distance	Combined	Property	Target	Entry	Optimal
Distance	--	$p \leq 0.05$	$p \leq 0.05$	$p \leq 0.01$	$p \leq 0.001$	*n.s.*
Combined		--	*n.s.*	*n.s.*	*n.s.*	$p \leq 0.05$
Property			--	*n.s.*	*n.s.*	$p \leq 0.05$
Target				--	*n.s.*	$p \leq 0.01$
Entry					--	$p \leq 0.001$
Optimal						--

6.8.2. Operating points as a measure of discrimination

As was the case in previous chapters, the ROC graphs in Figure 6.1 suggest that where the decision threshold is placed has a serious impact on discrimination accuracy. Youden's index was calculated in order to identify an optimal decision threshold for each behavioural domain. For inter-crime distances, the optimal threshold is $p \geq 0.01$ (≤ 2.0 km), whereas the optimal thresholds for the combined component, property stolen, target selection choices and entry behaviours are $p \geq 0.01$ (≥ 0.22), $p \geq 0.01$ (≥ 0.18), $p \geq 0.01$ (≥ 0.50), and $p \geq 0.01$ (≥ 0.20) respectively. One of the reasons the optimal decision thresholds are all so low (with respect to the p-values) is because of the large discrepancy between the number of unlinked and linked crime pairs in the Oldham residential sample.

6.8.3. Measuring improvements in discrimination accuracy

As in the previous chapter, when a variety of burglary behaviours are used to generate multiple ROC curves, it is of interest to determine the extent to which discrimination accuracy improves when using one behavioural domain over another. This issue was examined here by plotting the ROC curves and calculating how many more hits (and how many less false alarms) would be made at a particular false alarm rate (or hit rate) depending on the domain used.

This is illustrated in Table 6.9 using inter-crime distances as the ideal, a false alarm rate of 0.20 and a hit rate of 0.80.

As expected, more hits and less false alarms are made when inter-crime distances are used. At a fixed false alarm rate of 0.20, for example, 12 more hits can be made for every 100 crime pairs encountered if inter-crime distances are used instead of the combined component. In contrast, 47 more hits could be made if inter-crime distances were used instead of entry behaviours. Similarly, at a fixed hit rate of 0.80, 15 fewer false alarms can be made for every 100 crime pairs encountered if inter-crimes distances were used instead of the combined component. In contrast, 38 fewer false alarms could be made if inter-crime distances were used instead of entry behaviours.

Table 6.9. Improvements in discrimination accuracy

	Accuracy of inter-crime distances at:	
	*p*FA=0.20	*p*H=0.80
Compared to:		
Combined	+12 hits	-15 false alarms
Property	+28 hits	-18 false alarms
Target	+41 hits	-30 false alarms
Entry	+47 hits	-38 false alarms

6.9. The multiple feature ROC graph for residential burglary

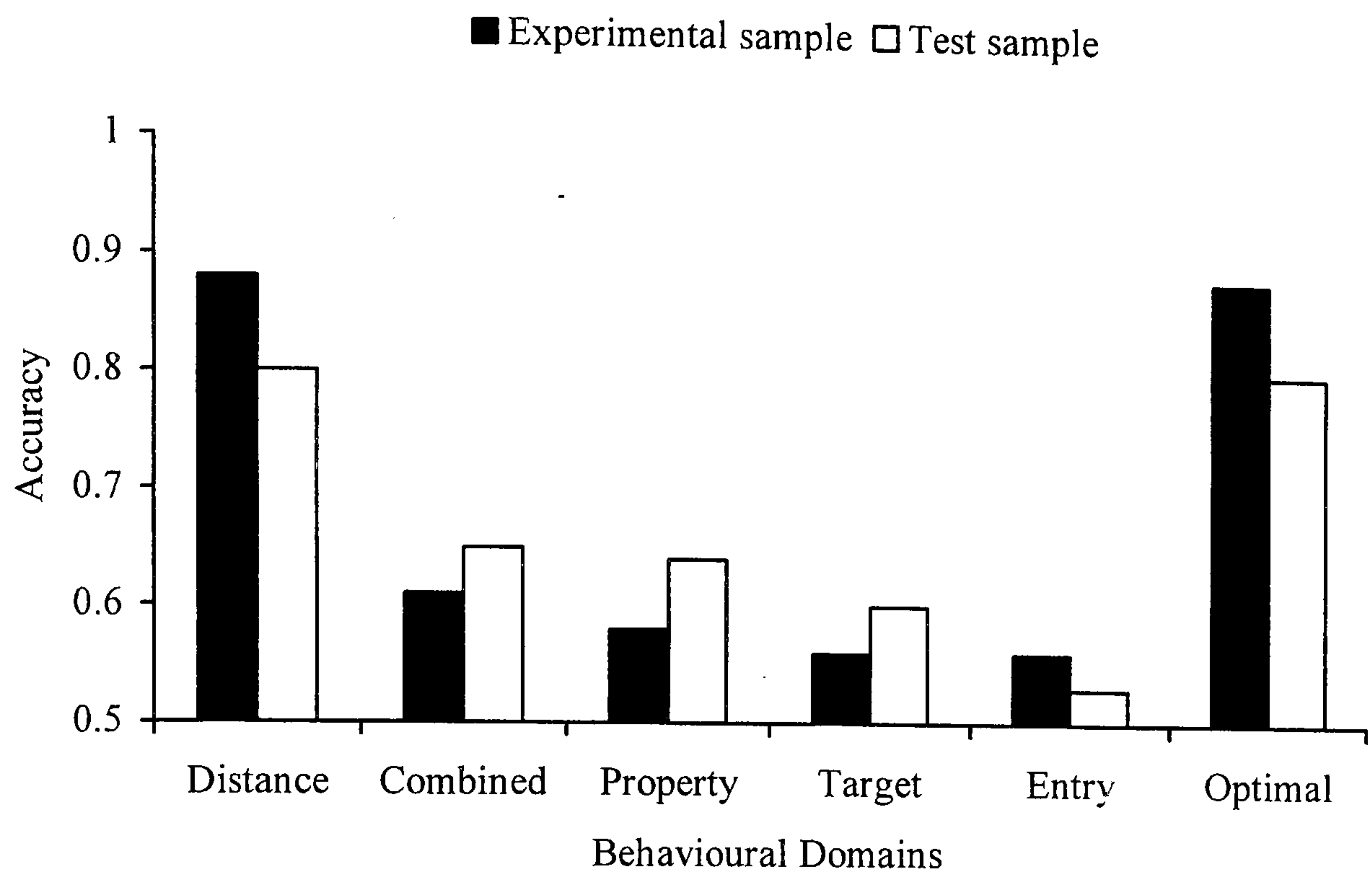
In order to get some indication of how valid the optimal regression model is for the Oldham residential burglary sample, it was also used to construct a ROC graph. This ROC graph, along with its AUC (and *p*-value), standard error, and 95% confidence interval, is presented as the bottom right ROC graph in Figure 6.1. Unsurprisingly, this figure is identical to the ROC graph for inter-crime distances, which is in line with the previous regression analysis.

6.10. Validating the empirical ROC curves for residential burglary

When the logistic regression models developed using data from the residential experimental sample are applied to data from the residential test sample, the ROC curves that result from the most accurate regression models have AUCs that are significantly greater than chance. This suggests that these logistic regression models do generalise to crimes beyond those used to construct the original regression models. In addition, the AUCs associated with each ROC curve generally correspond to how accurate the regression models are, with only target selection choices and entry behaviours switching places.

The validity of the logistic regression models can be tested more directly, however, by constructing ROC curves for each behavioural domain using data from the experimental sample, and comparing these to the ROC graphs in Figure 6.1. This is done in Figure 6.2 where it can be seen that the logistic regression models perform equally well using data from the test sample. Indeed, in relation to the combined component, property stolen, and target selection choices, the logistic regression models perform slightly better using data from the test sample.

Figure 6.2. Comparison of AUCs across the experimental and test samples



6.10.1. External discrimination trials

In addition to this validation procedure, a number of discrimination trials were carried out. Again, the goal was to determine whether the values of *p*H and *p*FA associated with points along the ROC curves in Figure 6.1 correspond to the values of *p*H and *p*FA obtained across discrimination trials. In this case, only the ROC curve associated with inter-crime distances was tested since this is the most effective discriminator for residential burglaries in Oldham. The values of *p*H and *p*FA at the optimal threshold of ≤2.00 km are 0.73 and 0.28 respectively. If the ROC graph provides a valid representation of how accurately discriminations can be made, it should be possible to achieve similar levels of accuracy across random discrimination trials when the same decision threshold is used.

Using the same method as in previous chapters, the results of these trials are presented in Table 6.10. Across 5 randomly selected samples of 1000 crime pairs, the average hit and false alarm rates were 0.82 and 0.26. Across 5 randomly selected samples of 10000 crime pairs, the average hit and false alarm rates were 0.70 and 0.19. Thus, the average hit and false alarm rates generally correspond to the predicted values regardless of sample size, though the average hit rate in the smaller samples is somewhat larger than the target value, and the average false alarm rate in the larger sample is somewhat smaller than the target value. Furthermore, on every trial there is a significant association between prediction and reality, as indicated by the X^2 values presented in the last column of the table.

Table 6.10. Validation for the Oldham residential burglary data

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.01$ ($\leq$2.00 km)	1000	0.92 (12)	0.08 (1)	0.74 (733)	0.26 (254)	29.13 (1)***
	$p \geq 0.01$ ($\leq$2.00 km)	10000	0.71 (416)	0.29 (173)	0.81 (7609)	0.19 (1802)	851.10 (1)***
2	$p \geq 0.01$ ($\leq$2.00 km)	1000	0.83 (10)	0.17 (2)	0.74 (731)	0.26 (257)	19.91 (1)***
	$p \geq 0.01$ ($\leq$2.00 km)	10000	0.69 (400)	0.31 (178)	0.81 (7643)	0.19 (1779)	809.25 (1)***
3	$p \geq 0.01$ ($\leq$2.00 km)	1000	1.00 (10)	0.00 (0)	0.74 (731)	0.26 (259)	27.45 (1)***
	$p \geq 0.01$ ($\leq$2.00 km)	10000	0.71 (414)	0.29 (169)	0.81 (7620)	0.19 (1797)	859.68 (1)***
4	$p \geq 0.01$ ($\leq$2.00 km)	1000	0.73 (11)	0.27 (4)	0.74 (727)	0.26 (258)	16.70 (1)***
	$p \geq 0.01$ ($\leq$2.00 km)	10000	0.70 (416)	0.30 (181)	0.81 (7596)	0.19 (1807)	826.92 (1)***
5	$p \geq 0.01$ ($\leq$2.00 km)	1000	0.64 (9)	0.36 (5)	0.72 (713)	0.28 (273)	9.13 (1)**
	$p \geq 0.01$ ($\leq$2.00 km)	10000	0.70 (412)	0.30 (173)	0.81 (7627)	0.19 (1788)	849.18 (1)***
Average	$p \geq 0.01$ ($\leq$2.00 km)	1000	0.82 (10.40)	0.18 (2.40)	0.74 (727.00)	0.26 (260.20)	--
	$p \geq 0.01$ ($\leq$2.00 km)	10000	0.70 (411.60)	0.30 (174.80)	0.81 (7619.00)	0.19 (1794.60)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

6.11. Single feature ROC graphs for commercial burglary

The single feature ROC graphs for commercial burglary, along with their AUCs (and *p*-values), standard errors, and 95% confidence intervals, are presented as the first 5 graphs in Figure 6.3. These ROC graphs correspond to the single feature regression models presented in Table 6.5 once they had been applied to each and every crime pair in the commercial test sample.

6.11.1. The AUC as a measure of spatial and behavioural consistency

Consistent with the analysis of data in the experimental sample, each of the ROC curves in Figure 6.3 indicate that commercial burglary behaviours in Oldham are expressed in a consistent fashion, though not all are consistent beyond what would be expected by chance. In addition, the ROC graphs confirm that certain behaviours are exhibited more consistently than others are. Again, inter-crime distances are the most consistent feature (AUC=0.82), followed by the combined component (AUC=0.68), entry behaviours (AUC=0.64), property stolen (AUC=0.59), and target selection choices (AUC=0.59). This ordering of the predictor variables is similar to the previous logistic regression analysis, although entry behaviours and target selection choices have switched places. Table 6.11 presents results showing which of these curves differ significantly from one another.

Figure 6.3. ROC graphs for Oldham commercial burglary data

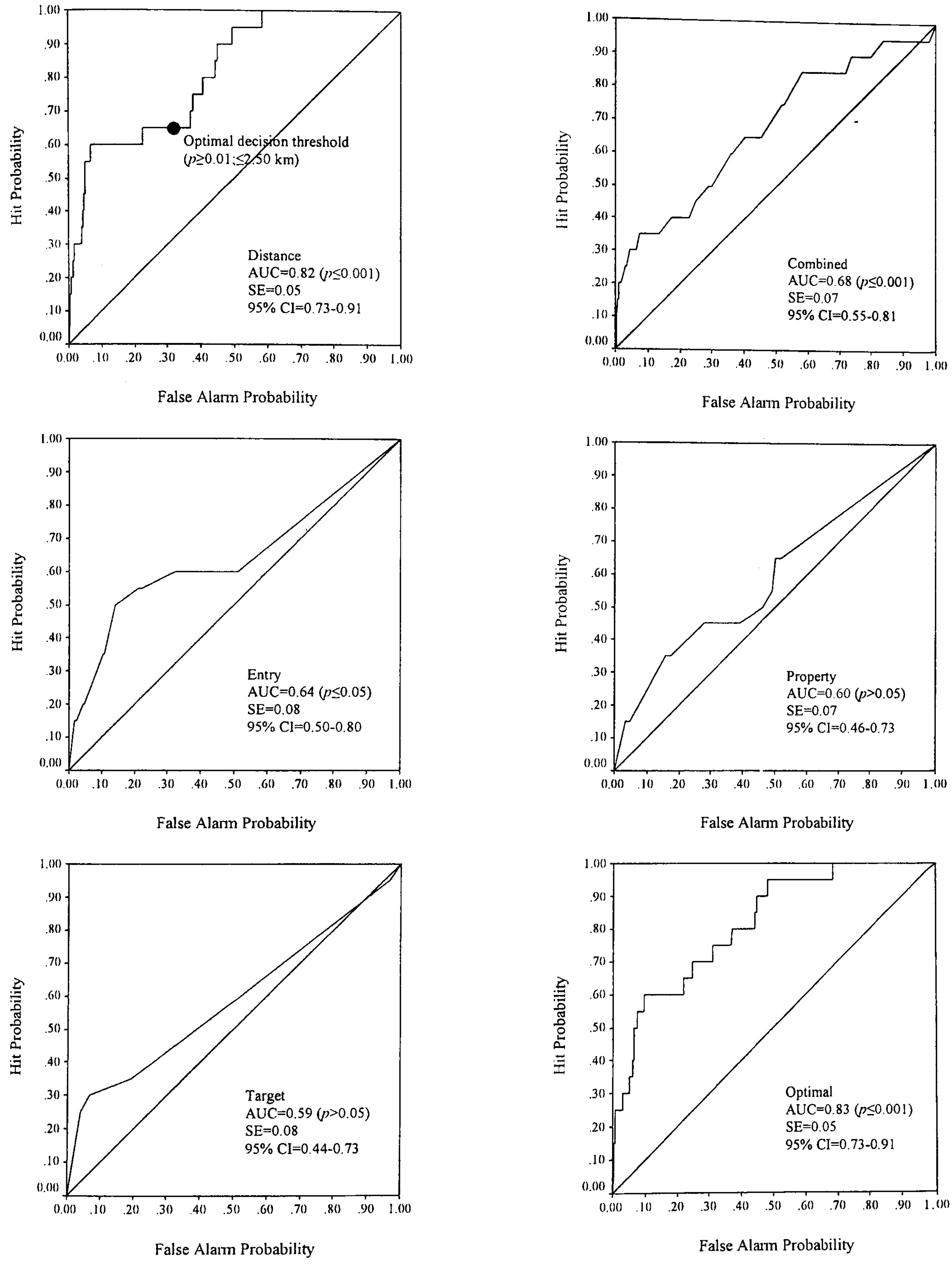


Table 6.11. Differences between the ROC curves in Figure 6.3

Variables	Distance	Combined	Entry	Property	Target	Optimal
Distance	--	$p \leq 0.05$	$p \leq 0.01$	$p \leq 0.01$	$p \leq 0.01$	*n.s.*
Combined		--	*n.s.*	*n.s.*	*n.s.*	$p \leq 0.05$
Entry			--	*n.s.*	*n.s.*	$p \leq 0.01$
Property				--	*n.s.*	$p \leq 0.01$
Target					--	$p \leq 0.01$
Optimal						--

6.11.2. Operating points as a measure of discrimination

The ROC graphs presented in Figure 6.3 again suggest that the decision threshold has a serious impact on discrimination accuracy. Youden's index was calculated in order to identify an optimal decision threshold for each behavioural domain. For inter-crime distances, the optimal threshold is $p \geq 0.01$ (≤ 2.50 km), whereas the optimal thresholds for the combined component, entry behaviours, property stolen, and target selection choices are $p \geq 0.01$ (≥ 0.28), $p \geq 0.01$ (≥ 0.28), $p \geq 0.01$ (≥ 0.35), and $p \geq 0.01$ (≥ 0.48) respectively. Again, one of the reasons why the optimal decision thresholds are all so low (with respect to the p-values) is because of the large discrepancy between the number of unlinked and linked crime pairs in the Oldham commercial sample.

6.11.3. Measuring improvements in discrimination accuracy

As before, the extent to which discrimination accuracy improves when one behavioural domain is used over another was examined in relation to commercial burglary. The results are illustrated in Table 6.12, using inter-crime distances as the ideal, a false alarm rate of 0.20, and a hit rate of 0.80. As expected, more hits and fewer false alarms are made when inter-crime distances are used. For example, at a fixed false alarm rate of 0.20, 20 more hits can be made for every

100 crime pairs encountered if inter-crime distances are used as the basis for making decisions instead of the combined component. In contrast, 25 more hits could be made if inter-crime distances were used instead of target selection choices. Similarly, at a fixed hit rate of 0.80, 10 fewer false alarms can be made for every 100 crime pairs encountered if inter-crimes distances were used instead of the combined component. In contrast, 32 fewer false alarms could be made if inter-crime distances were used instead of target selection choices.

Table 6.12. Improvements in discrimination accuracy

	Accuracy of inter-crime distances at:	
	*p*FA=0.20	*p*H=0.80
Compared to:		
Combined	+20 hits	-10 false alarms
Entry	+5 hits	-30 false alarms
Property	+23 hits	-28 false alarms
Target	+25 hits	-32 false alarms

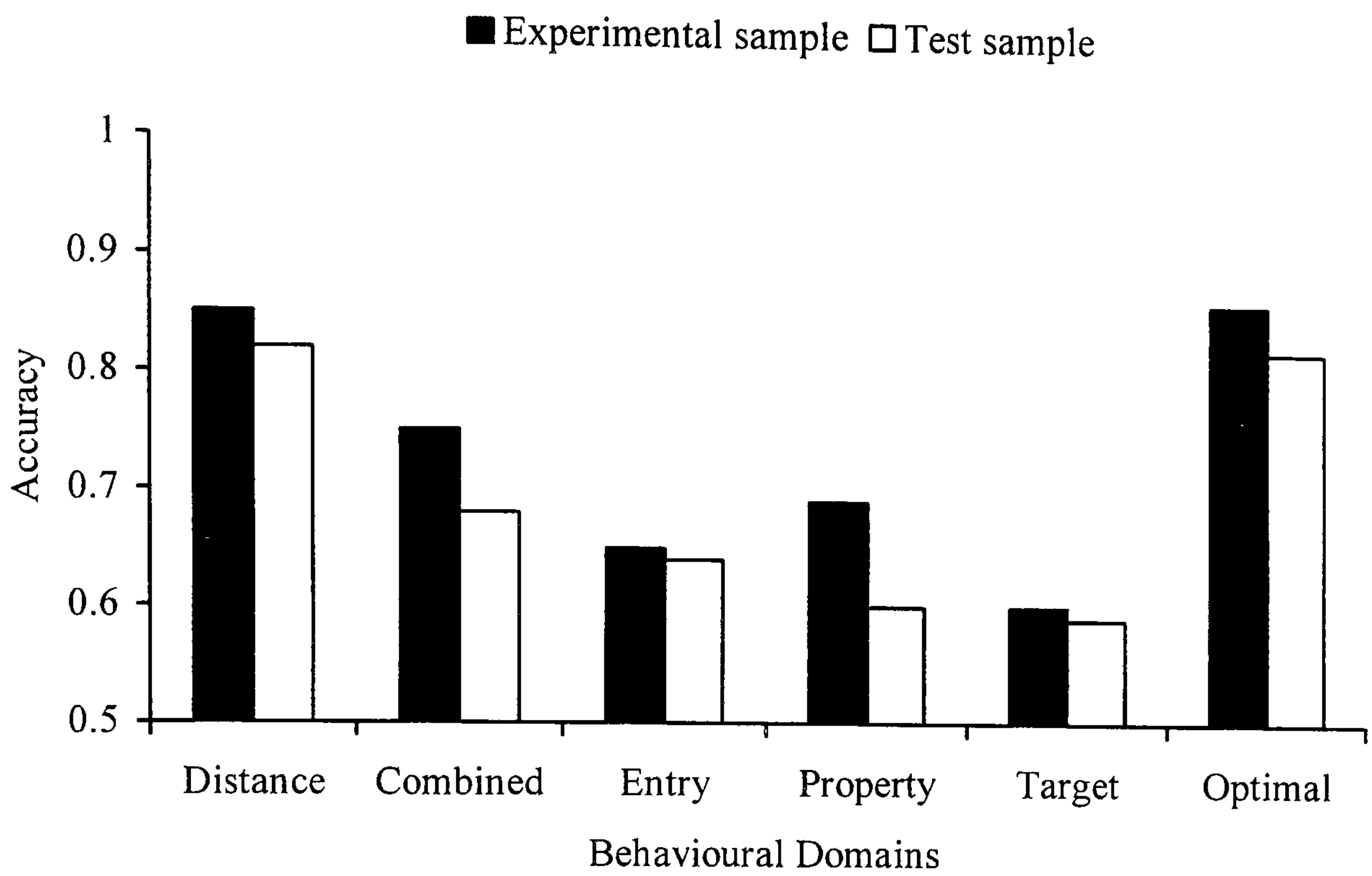
6.12. The multiple feature ROC graph for commercial burglary

In order to get some indication of how valid the optimal regression model is for commercial burglary behaviours, it was also used to construct a ROC graph. This ROC graph, along with its AUC (and *p*-value), standard error, and 95% confidence interval, is presented as the bottom right ROC graph in Figure 6.3. As with the other ROC graphs in Figure 6.3, this ROC graph supports the previous regression analysis. Not only does the optimal regression model result in a ROC curve that is significantly more accurate than chance (AUC=0.83; $p \leq 0.001$), it also results in a ROC curve that is slightly more accurate than any of the single feature ROC curves.

6.13. Validating the empirical ROC curves for commercial burglary

When the logistic regression models developed using data from the commercial experimental sample are applied to data from the commercial test sample, the ROC curves that result from the most accurate regression models have AUCs that are significantly greater than chance. This suggests that these logistic regression models do generalise to crimes beyond those used to construct the original models. In addition, the AUCs associated with each ROC curve generally correspond to how accurate the regression models are, with only entry behaviours and target selection choices switching places. The validity of the logistic regression models was tested more directly by constructing ROC curves for each behavioural domain using data from the experimental sample, and comparing these to the ROC graphs in Figure 6.3. This is done in Figure 6.4 where it can be seen that the logistic regression models perform almost as well using data from the test sample.

Figure 6.4. Comparison of AUCs across experimental and test samples



6.13.1. External discrimination trials

In addition to this validation procedure, a number of discrimination trials were carried out. Again, only the ROC curve associated with inter-crime distances was tested since this is the most effective discriminator for commercial burglaries in Oldham. Recall that the values of *p*H and *p*FA achieved for the optimal decision threshold of ≤2.50 km are 0.65 and 0.38 respectively. If the ROC graph provides a valid representation of how accurately discriminations can be made, it should be possible to achieve similar levels of accuracy across random discrimination trials when the same decision threshold is used.

Using the same method as before, the results of these trials are presented in Table 6.13. Across 5 randomly selected samples of 1000 crime pairs, the average hit and false alarm rates were 0.72 and 0.32. Across 5 randomly selected samples of 7000 crime pairs, the average hit and false alarm rates were 0.87 and 0.28. Thus, the average hit and false alarm rates from the smaller discrimination trials appear to correspond to the target values, though the rates generated from the larger samples are not quite as close. However, on every trial (excluding trial 5 for the smaller sample) there is a highly significant association between prediction and reality, as indicated by the X^2 values presented in the last column of the table.

Table 6.13. Validation trials for Oldham commercial burglary data

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.01$ ($\leq$2.50 km)	1000	0.80 (12)	0.20 (3)	0.69 (678)	0.31 (307)	16.22 (1)***
	$p \geq 0.01$ ($\leq$2.50 km)	7000	0.87 (133)	0.13 (20)	0.73 (4974)	0.27 (1873)	259.88 (1)***
2	$p \geq 0.01$ ($\leq$2.50 km)	1000	0.75 (12)	0.25 (4)	0.69 (677)	0.31 (307)	13.90 (1)***
	$p \geq 0.01$ ($\leq$2.50 km)	7000	0.87 (135)	0.13 (20)	0.72 (4935)	0.28 (1910)	256.81 (1)***
3	$p \geq 0.01$ ($\leq$2.50 km)	1000	0.77 (10)	0.23 (3)	0.68 (668)	0.32 (319)	11.56 (1)***
	$p \geq 0.01$ ($\leq$2.50 km)	7000	0.82 (130)	0.18 (18)	0.72 (4938)	0.28 (1914)	251.47 (1)***
4	$p \geq 0.01$ ($\leq$2.50 km)	1000	0.63 (10)	0.37 (6)	0.68 (674)	0.32 (310)	6.95 (1)**
	$p \geq 0.01$ ($\leq$2.50 km)	7000	0.89 (133)	0.11 (17)	0.72 (4966)	0.28 (1884)	267.71 (1)***
5	$p \geq 0.01$ ($\leq$2.50 km)	1000	0.63 (5)	0.37 (3)	0.67 (669)	0.33 (323)	3.23 (1)
	$p \geq 0.01$ ($\leq$2.50 km)	7000	0.88 (143)	0.12 (19)	0.72 (4926)	0.28 (1912)	277.55 (1)***
Average	$p \geq 0.01$ ($\leq$2.50 km)	1000	0.72 (9.80)	0.28 (3.80)	0.68 (673.20)	0.32 (313.20)	--
	$p \geq 0.01$ ($\leq$2.50 km)	7000	0.87 (124.80)	0.13 (18.80)	0.72 (4947.80)	0.28 (1898.60)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

6.14. Chapter summary

In this chapter, the behaviour of residential and commercial burglars who committed crimes in Oldham was explored. This extended the analysis of burglary behaviour presented in the previous chapters in two important ways. First, the burglaries examined in this chapter were selected from a single police division rather than an entire police force, as was the case with London and Dorset. Second, a range of behaviours exhibited by commercial burglars was examined in this chapter in addition to the various residential burglary behaviours that were examined in London and Dorset.

Descriptive statistics provided partial support for the hypothesis that offenders in Oldham exhibit burglary behaviours in a consistent fashion. In the case of residential burglary, significant differences were found between the similarity scores for unlinked and linked crime pairs in relation to inter-crime distances, the combined component and property stolen, with all the differences being in the expected direction. In the case of commercial burglary, significant differences were found for these three behavioural domains as well as for entry behaviours and target selection choices, with all the differences being in the expected direction.

Logistic regression analysis was run on the similarity scores associated with each behavioural domain as a way of calculating predicted probabilities that crime pairs were linked. These probabilities, in turn, formed the basis of ROC analysis. Consistent with the descriptive analysis, ROC analysis indicated that inter-crime distances are the most consistent aspect of both residential and commercial burglary behaviour in Oldham. In line with this, relatively high levels of discrimination accuracy could also be achieved when using these inter-crime distances.

When using a distance threshold of ≤ 2.00 for residential burglaries in Oldham, 73% of linked burglaries were correctly classified as linked and 28% of unlinked burglaries were incorrectly classified as linked. When using a distance threshold of ≤ 2.50 for commercial burglaries in Oldham, 65% of linked burglaries were correctly classified as linked and 38% of unlinked burglaries were incorrectly classified as linked. Similar results emerged across discrimination trials, suggesting the ROC procedure does have an adequate level of predictive accuracy in Oldham.

As was the case in the previous two chapters, these results confirm that residential burglars are spatially consistent with respect to their inter-crime distances, and that these distances result in higher levels of discrimination accuracy compared to any of the other crime scene behaviours. However, the results presented in this chapter extend these previous findings in a variety of ways. First, the results indicate that commercial burglars in Oldham also choose their crime sites in a relatively consistent fashion, and that this aspect of burglary is exhibited more consistently than any other aspect. Second, the results indicate that inter-crime distances in commercial burglary result in higher levels of discrimination accuracy compared to any of the other crime scene behaviours. Third, the results indicate that high levels of discrimination accuracy can be achieved in cases of residential and commercial burglary even when the burglaries have been sampled at a divisional level.

These new findings are extremely important because they imply that a similar linking strategy (i.e., basing linking decisions on inter-crime distances) has the potential to be productively utilised when dealing with different types of burglary in a single police division. This may be viewed as surprising considering that differences between residential and commercial burglars are expected with respect to certain aspects of their spatial behaviour (Capone & Nichols, 1976; Van Koppen & Jansen, 1998). Indeed, such differences were found in Oldham as well, with residential burglars exhibiting slightly shorter inter-crime distances compared to commercial burglars. What is indicated by the results in the present chapter, however, is that even when such differences do exist, relatively high levels of discrimination accuracy can be achieved for both types of crime so long as appropriate decision thresholds are identified and adopted.

CHAPTER 7
THE BEHAVIOUR OF SERIAL BURGLARS IN MERSEYSIDE

7.1. Introduction

In contrast to previous chapters, where a single sample of burglars was examined from each of the participating police forces, this chapter explores the behaviour of residential and commercial burglars selected from the 4 different police districts in Merseyside. As was the case with Oldham burglaries, the data collected from Merseyside also allows a variety of different behaviours to be examined for both types of crime. Once again, the objectives in this chapter are to determine whether various aspects of burglary behaviour are expressed in a consistent fashion by burglars in the Merseyside districts, and if so, whether an analysis of these behaviours can form a reliable basis for distinguishing between offenders.

The analysis presented in this chapter will give some indication as to whether it is necessary to develop linking strategies for burglary at a local level, or whether a general strategy could be as effective. While it appears that a more general strategy is the norm in cases of serial rape and murder (e.g., Grubin *et al.*, 2001; Keppel & Weis, 1993), this issue is not something that has ever been explored in relation to serial burglary. Unlike serial rape and murder, the extremely local nature of most burglaries, coupled with the enormous number of burglaries that are committed, suggests that a reasonably small-scale approach would be the only effective option when dealing with serial burglary. The feasibility of developing such a strategy, however, depends on whether high levels of consistency and discrimination can be found in burglary behaviour at the local level.

7.1.1. The area

The residential burglaries included in the Merseyside sample were committed between January 1995 and December 1999 across the Merseyside police districts, while the commercial burglaries were committed between January 1994 and December 1999. Merseyside covers an area of 655 km^2, has a population of about one and a half million persons, and a population density of approximately 2143 persons/km^2.

The Merseyside Police Service consists of 4081 police officers. These officers patrol the 4 different police districts making up Merseyside. These include the Wirral district, the Sefton district, the Liverpool district and the St. Helen's and Knowsley district (see Figure 7.1). These districts differ in terms of their demographic and topographic characteristics. To summarise some of these differences, Table 7.1 includes information about the areas covered by each police district, the current populations and population densities, and the number of residential burglaries recorded and cleared by the police during the year 2001. Unfortunately, similar information for commercial burglaries is not readily available. What the information in Table 7.1 makes clear is that despite the many differences that exist between police districts in Merseyside, the crime of burglary represents a very serious problem that needs to be addressed.

Figure 7.1. The police districts in Merseyside

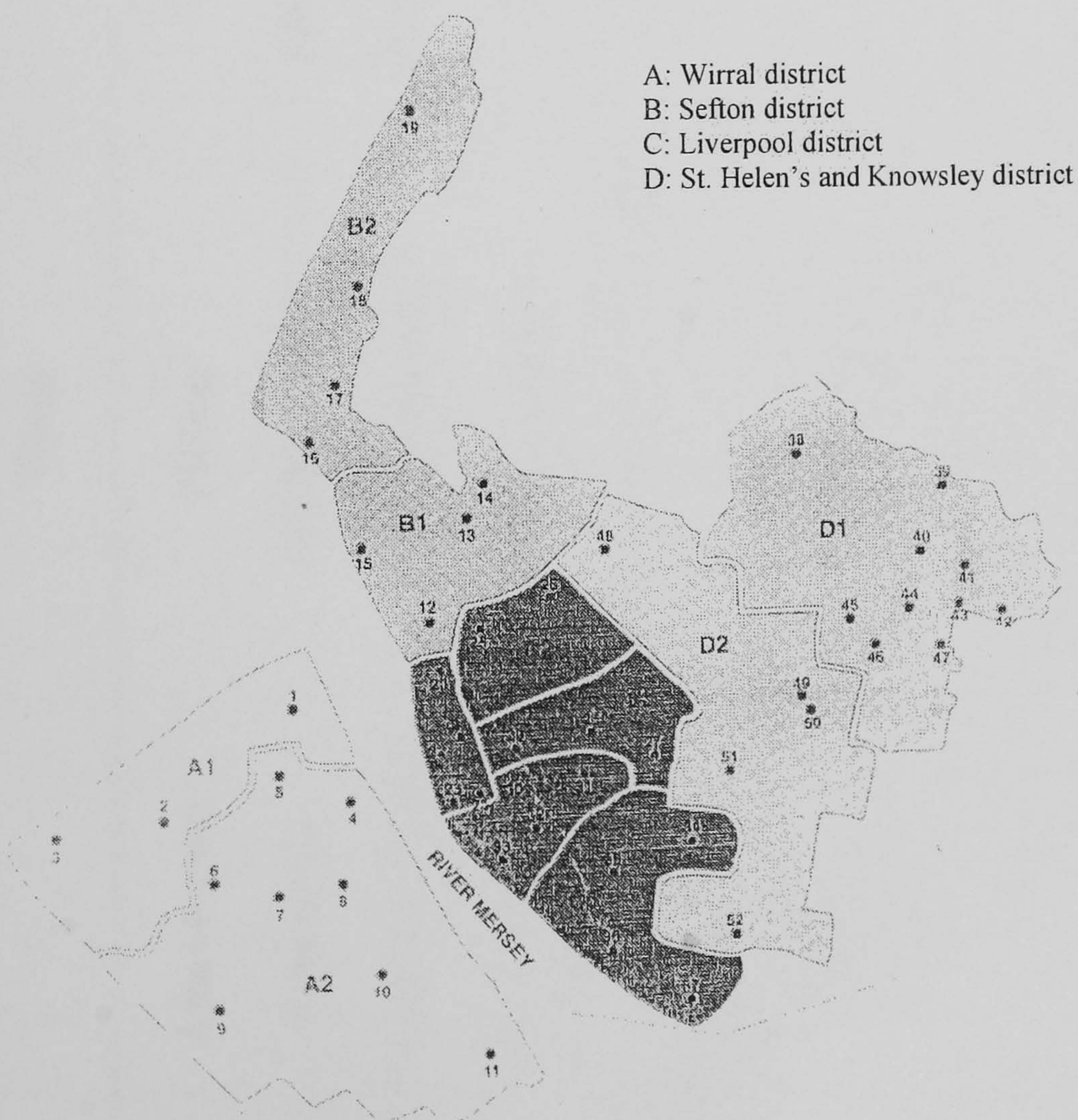


Table 7.1. A brief description of the police districts in Merseyside

	A Wirral	B Sefton	C Liverpool	D St. Helen’s and Knowsley
Area covered	159 km^2	153 km^2	113 km^2	230 km^2
Population	327000 persons	286000 persons	457000 persons	439000 persons
Population density	2054 persons/km^2	1875 persons/km^2	4053 persons/km^2	1467 persons/km^2
Residential burglary (reported/cleared)	2852 / 456	1528 / 244	5411 / 866	2723 / 436

7.1.2. The data

As was the case with the data collected from London, Dorset and Oldham, the information pertaining to spatial behaviour consists of geo-coded x and y coordinates attached to each burglary location. For all other behaviours, the information was coded in dichotomous form, with a 1 indicating that a particular behaviour was present and a 0 indicating that it was absent. A more detailed list of the behaviours included in the residential and commercial burglary samples can be found in Appendix H and I.

The entire residential burglary sample collected from Merseyside consists of 51 serial burglars responsible for 660 crimes. The offence series range in length from 3 crimes to 23 crimes. For the purpose of the analysis presented in this chapter, however, a smaller subset of residential burglaries was selected from this sample. This reduced sample was broken down further into 4 district samples. These samples will henceforth be referred to as Merseyside residential A, B, C and D, which correspond to the Wirral district, the Sefton district, the Liverpool district, and the St. Helen's and Knowsley district respectively.

The entire commercial burglary sample collected from Merseyside consists of 57 serial burglars responsible for 634 crimes. The offence series range in length from 3 crimes to 24 crimes. For the purpose of the analysis presented in this chapter, however, a smaller subset of commercial burglaries was also selected from this sample. This reduced sample was broken down further into 4 district samples. These samples will henceforth be referred to as Merseyside commercial A, B, C and D, which correspond to the Wirral district, the Sefton district, the Liverpool district, and the St. Helen's and Knowsley district respectively.

7.2. Calculating spatial and behavioural similarity scores

In order to examine the behaviour of residential and commercial burglars committing crimes in Merseyside, spatial and behavioural similarity scores needed to be calculated. These scores were calculated separately for each residential and commercial district sample. As before, spatial similarity scores consist of inter-crime distances, and were calculated by entering the x and y coordinates from each district sample into *S-LINK*. The behavioural similarity scores consist of Jaccard coefficients, calculated for each behavioural domain by entering the dichotomous data from each district sample into *B-LINK*. All generated crime pairs from *S-LINK* and *B-LINK* were then defined as unlinked or linked, based on who was known to have committed the crimes. The number of crime pairs resulting from this analysis is summarised in Table 7.2.

Table 7.2. Number of crime pairs per district

	Residential		Commercial	
	Unlinked crime pairs	Linked crime pairs	Unlinked crime pairs	Linked crime pairs
A	669	72	55	23
B	1225	101	2430	126
C	9322	269	9333	258
D	250	50	1713	117

7.3. A descriptive analysis of the spatial and behavioural similarity scores

As in previous chapters, the first step in examining issues of consistency and discrimination was to calculate descriptive statistics. These were calculated across all unlinked and linked crime pairs for each district separately. As before, it was expected that similarity scores calculated across unlinked crime pairs would be low relative to similarity scores calculated across linked crime pairs. If this were found to be the case, it would provide support that burglary behaviours in Merseyside are expressed consistently over time, thus making it possible to discriminate between crimes committed by different offenders.

7.3.1. A descriptive analysis of residential burglary behaviours

The descriptive analysis of spatial and behavioural similarity scores for residential burglary is presented in Table 7.3. This table includes the mean values of similarity scores, along with their ranges and standard deviations. In addition, the results from t-tests are provided. As indicated by the mean values, unlinked crime pairs in residential burglary consistently have lower similarity scores compared to linked crime pairs across each police district. However, cases exist where high levels of similarity are found for unlinked crime pairs, as well as low levels of similarity for linked crime pairs. Despite this, the t-tests presented in Table 7.3 indicate that the vast majority of differences are highly significant, excluding most of the similarity scores based on internal behaviours.

Table 7.3. Summary of the Merseyside residential burglary data

		Unlinked crime pairs			Linked crime pairs			
	District	M	Range	SD	M	Range	SD	t
Distance	A	3.36	0.13-9.96	2.21	1.50	0.00-8.66	1.63	6.92***
	B	10.73	0.15-23.88	7.00	1.54	0.01-5.92	1.23	13.18***
	C	4.71	0.02-16.34	2.77	1.23	0.00-5.15	1.25	20.59***
	D	7.81	0.00-13.44	4.84	1.53	0.00-5.72	1.60	9.01***
Combined	A	0.25	0.00-0.75	0.13	0.36	0.08-1.00	0.17	4.77***
	B	0.25	0.05-1.00	0.18	0.31	0.07-1.00	0.20	3.16**
	C	0.25	0.00-1.00	0.15	0.38	0.00-1.00	0.22	13.98***
	D	0.22	0.00-0.88	0.14	0.30	0.07-0.70	0.16	3.59***
Entry	A	0.22	0.00-1.00	0.22	0.30	0.00-1.00	0.25	2.60**
	B	0.13	0.00-1.00	0.19	0.20	0.00-1.00	0.24	3.46***
	C	0.21	0.00-1.00	0.24	0.35	0.00-1.00	0.31	9.79***
	D	0.23	0.00-1.00	0.22	0.33	0.00-1.00	0.33	2.81**
Property	A	0.01	0.00-1.00	0.17	0.13	0.00-1.00	0.26	2.62**
	B	0.04	0.00-1.00	0.14	0.08	0.00-1.00	0.19	3.01**
	C	0.08	0.00-1.00	0.17	0.14	0.00-1.00	0.21	5.52***
	D	0.07	0.00-1.00	0.15	0.12	0.00-0.50	0.16	2.21*
Target	A	0.62	0.00-1.00	0.33	0.74	0.00-1.00	0.32	2.92**
	B	0.64	0.33-1.00	0.33	0.68	0.33-1.00	0.33	1.05
	C	0.63	0.00-1.00	0.34	0.73	0.00-1.00	0.33	4.72***
	D	0.49	0.00-1.00	0.33	0.66	0.33-1.00	0.32	3.35***
Internal	A	0.01	0.00-1.00	0.07	0.01	0.00-1.00	0.05	0.02
	B	0.03	0.00-1.00	0.14	0.05	0.00-1.00	0.21	1.61
	C	0.01	0.00-1.00	0.07	0.02	0.00-1.00	0.11	2.50**
	D	0.00	0.00-0.50	0.04	0.01	0.00-0.50	0.07	1.05

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

7.3.2. A descriptive analysis of commercial burglary behaviours

The descriptive analysis of spatial and behavioural similarity scores for commercial burglary is presented in Table 7.4. This table includes the mean values of similarity scores, along with their ranges and standard deviations. In addition, the results from t-tests are provided. As indicated by the mean values, unlinked crime pairs in commercial burglary also have consistently lower similarity scores compared to linked crime pairs across each police district. Furthermore, as in all previous analyses, cases exist where high levels of similarity are found for unlinked crime pairs, as well as low levels of similarity for linked crime pairs. Despite this, the t-tests again indicate that the vast majority of differences between similarity scores are highly significant, excluding many of the comparisons in commercial A. To some extent, however, the lack of significant findings in commercial A is probably due to a small sample size.

Table 7.4. Summary of the Merseyside commercial burglary data

		Unlinked crime pairs			Linked crime pairs			
	District	M	Range	SD	M	Range	SD	t
Distance	A	6.32	0.84-11.88	3.58	4.48	0.00-9.54	3.15	2.14*
	B	7.45	0.00-26.12	7.34	3.31	0.00-23.19	5.60	6.24***
	C	4.76	0.00-16.13	3.40	1.15	0.00-6.32	1.32	17.01***
	D	8.77	0.00-19.73	4.77	1.31	0.00-6.94	1.34	16.87***
Combined	A	0.08	0.00-0.38	0.10	0.11	0.00-0.36	0.12	0.84
	B	0.13	0.00-1.00	0.14	0.20	0.00-0.80	0.19	5.40***
	C	0.11	0.00-1.00	0.11	0.18	0.00-0.82	0.17	10.72***
	D	0.14	0.00-0.88	0.14	0.27	0.00-1.00	0.24	9.16***
Target	A	0.01	0.00-1.00	0.13	0.09	0.00-1.00	0.29	1.44
	B	0.08	0.00-1.00	0.28	0.25	0.00-1.00	0.44	6.54***
	C	0.06	0.00-1.00	0.24	0.29	0.00-1.00	0.46	14.60***
	D	0.07	0.00-1.00	0.26	0.28	0.00-1.00	0.45	8.06***
Property	A	0.05	0.00-1.00	0.17	0.13	0.00-1.00	0.34	1.44
	B	0.04	0.00-1.00	0.17	0.11	0.00-1.00	0.30	3.98***
	C	0.05	0.00-1.00	0.18	0.09	0.00-1.00	0.24	3.28***
	D	0.06	0.00-1.00	0.22	0.27	0.00-1.00	0.43	9.18***
Entry	A	0.14	0.00-1.00	0.20	0.11	0.00-1.00	0.16	0.56
	B	0.20	0.00-1.00	0.22	0.25	0.00-1.00	0.27	2.70**
	C	0.17	0.00-1.00	0.21	0.24	0.00-1.00	0.28	5.29***
	D	0.20	0.00-1.00	0.19	0.28	0.00-1.00	0.25	4.55***
Internal	A	0.00	0.00-0.00	0.00	0.06	0.00-0.50	0.16	2.79**
	B	0.07	0.00-1.00	0.21	0.10	0.00-1.00	0.26	1.83
	C	0.05	0.00-1.00	0.17	0.09	0.00-1.00	0.23	3.00**
	D	0.11	0.00-1.00	0.29	0.21	0.00-1.00	0.39	3.42***

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

7.4. Logistic regression analysis

The descriptive analysis presented in the previous section provides support that burglars committing crimes in the various Merseyside districts express a range of offence behaviours in a consistent fashion. This, in turn, suggests it may be possible to use these behaviours to discriminate between crimes committed by different offenders. However, as in previous chapters, the fact that there is a degree of overlap between unlinked and linked similarity scores suggests that the level of consistency exhibited by Merseyside burglars will not be absolute, and that discrimination accuracy will not be perfect. Furthermore, the fact that the degree of overlap varies depending on the domain and police district considered suggests that consistency and discrimination will be domain and district specific.

To examine these issues more directly, single feature and optimal feature logistic regression models were developed from the residential and commercial burglary data obtained from each district sample. Again, the development of these models takes the analysis of burglary behaviour beyond a simple examination of average across-crime similarity scores. Each model provides further evidence of the degree to which residential and commercial burglaries committed in the various police districts of Merseyside can be discriminated from one another, and each models provides the necessary data for ROC analysis.

7.4.1. Validation datasets

As in previous chapters, each district sample was split randomly in half for the purpose of validating the logistic regression models. The logistic regression models were developed on the experimental samples and tested for generalisation on the test samples. Tables 7.5 (a, b, c and d) and 7.7 (a, b, c and d) contain a summary of the logistic regression models. As before, a range of information is provided in these tables, including the model coefficients and standard errors (constant and logit), an indicator of predictive accuracy for each predictor variable (Wald's statistic), and indices of general model fit (R^2 and X^2).

7.5. Logistic regression models for residential burglary

Each of the single feature regression models constructed from the residential district samples will be discussed first before moving on to the optimal models.

All of these single feature models were constructed using direct logistic regression analysis, and include a model developed for inter-crime distances, the combined component, entry behaviours, target selection choices, property stolen and internal behaviours.

Table 7.5 (a). Logistic regression models for residential A

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-0.74 (0.30)	-0.70 (0.16)	19.06 (1)***	0.18	32.89 (1)***
Combined	-3.72 (0.44)	5.12 (1.22)	17.66 (1)***	0.10	18.62 (1)***
Entry	-2.72 (0.28)	1.93 (0.73)	7.00 (1)**	0.04	6.56 (1)**
Property	-2.38 (0.20)	1.59 (0.78)	4.17 (1)*	0.02	3.62 (1)
Target	-2.84 (0.42)	0.90 (0.54)	2.79 (1)	0.02	2.85 (1)
Internal	-2.24 (0.18)	1.48 (1.87)	0.63 (1)	0.00	0.51 (1)
Optimal	-0.70 (0.31)	--	--	0.19	35.00 (2)***
Distance	--	-0.74 (0.17)	19.45 (1)***	--	--
Internal	--	3.47 (2.01)	2.98 (1)	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 7.5 (b). Logistic regression models for residential B

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	0.32 (0.30)	-0.77 (0.13)	37.39 (1)***	0.44	135.22 (1)***
Entry	-2.77 (0.19)	1.77 (0.64)	7.69 (1)**	0.03	6.92 (1)**
Property	-2.59 (0.16)	1.60 (0.69)	5.39 (1)*	0.02	4.48 (1)*
Combined	-2.76 (0.24)	1.00 (0.68)	2.19 (1)	0.01	1.98 (1)
Target	-2.67 (0.33)	0.28 (0.44)	0.41 (1)	0.00	0.41 (1)
Internal	-2.48 (0.15)	0.26 (1.02)	0.06 (1)	0.00	0.07 (1)
Optimal	0.03 (0.32)	--	--	0.49	152.40 (2)***
Distance	--	-0.87 (0.14)	39.31 (1)***	--	--
Entry	--	3.63 (0.91)	16.05 (1)***	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 7.5 (c). Logistic regression models for residential C

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-0.74 (0.15)	-1.14 (0.09)	163.96 (1)***	0.31	343.66 (1)***
Combined	-4.74 (0.17)	3.90 (0.38)	105.76 (1)***	0.08	86.50 (1)***
Entry	-4.14 (0.14)	2.18 (0.30)	54.46 (1)***	0.05	49.21 (1)***
Property	-3.71 (0.10)	1.54 (0.37)	16.94 (1)***	0.01	13.76 (1)***
Target	-4.18 (0.21)	0.94 (0.26)	12.69 (1)***	0.01	13.13 (1)***
Internal	-3.55 (0.08)	1.28 (0.73)	3.10 (1)	0.00	2.26 (1)
Optimal	-1.46 (0.20)	--	--	0.34	385.22 (4)***
Distance	--	-1.09 (0.09)	155.13 (1)***	--	--
Entry	--	1.65 (0.34)	23.95 (1)***	--	--
Property	--	1.68 (0.46)	13.17 (1)***	--	--
Internal	--	1.71 (0.20)	3.76 (1)*	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 7.5 (d). Logistic regression models for residential D

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	0.04 (0.33)	-0.46 (0.11)	16.01 (1)***	0.41	41.39 (1)***
Target	-2.39 (0.46)	1.42 (0.66)	4.60 (1)*	0.05	4.58 (1)*
Combined	-2.32 (0.45)	3.00 (1.52)	3.90 (1)*	0.04	3.79 (1)*
Entry	-2.01 (0.33)	1.44 (0.80)	3.20 (1)	0.03	3.08 (1)
Property	-1.74 (0.25)	1.91 (1.56)	1.49 (1)	0.02	1.39 (1)
Internal	-1.60 (0.22)	-18.39 (88.96)	0.04 (1)	0.00	0.36 (1)
Optimal	0.04 (0.33)	--	--	0.41	41.39 (1)***
Distance	--	-0.45 (0.11)	16.01 (1)***	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

7.5.1. The regression coefficients

Consistent with the findings presented in Table 7.3 the signs of the logit coefficients in the single feature models indicate that unlinked crime pairs tend to be characterised by larger inter-crime distances and lower levels of across-crime similarity than linked crime pairs. There is only one case where this is not true, for the similarity scores associated with internal behaviours in residential D. However, the standard error associated with the regression coefficient in this case suggests there is a problem, probably due to the relatively small sample size.

As before, to determine what these logit coefficients mean, each of them was multiplied by 0.10 and exponentiated. The results from this analysis are presented in Table 7.6 where it can be seen that an increase of 0.10 units for inter-crime distances would cause a decrease in the odds of two crimes being linked. In contrast, an increase of 0.10 units for all other similarity scores would cause an increase in the odds of two crimes being linked, with the exception of internal behaviours in residential D.

Table 7.6. The odds of crimes being linked as a function of similarity

	District			
Variables	A	B	C	D
Distance	0.93	0.93	0.89	0.96
Combined	1.11	1.67	1.48	1.35
Entry	1.19	1.21	1.24	1.15
Property	1.17	1.17	1.17	1.21
Target	1.03	1.09	1.09	1.15
Internal	1.03	1.16	1.14	0.16[a]

[a] Contrary to prediction

The impact of changes in the predictor variables was also assessed by examining changes in the probability that two crimes are linked. Unsurprisingly, considering the odds calculations just carried out, the general effect of increasing across-crime similarity scores in residential burglary is an increase in the probability that two crimes are linked. However, the rate of increase was domain specific.

7.5.2. Predictive accuracy and goodness-of-fit

Also consistent with the findings presented in Table 7.3, many of the logistic regression models constructed from data in the residential district samples were found to have a high degree of predictive accuracy and fit. A number of consistencies emerge from these measures. First, the regression models including inter-crime distances are always characterised by the highest levels of accuracy and fit. Second, the regression models including internal behaviours are always characterised by the lowest levels of accuracy and fit. There are no other consistent patterns of accuracy and fit associated with the other behavioural domains, indicating that if one of these domains is consistent in a particular police district this does not necessarily mean that the same domain will be consistent in another district.

7.5.3. The multiple feature models

In addition to constructing single feature logistic regression models from the residential samples, optimal regression models were constructed (the combined component was again not included in these analyses). Summaries of the optimal logistic regression models are also presented in Table 7.5 (a, b, c and d). What can be seen from these tables is that the behavioural domains included in the optimal models are dependent on the police district considered, as would be expected from the single feature models.

In residential A, for example, the optimal model includes 2 of the 5 predictor variables (inter-crime distances and internal behaviours). In residential B, the optimal model also includes 2 of the 5 predictor variables (inter-crime distances and entry behaviours). In residential C, the optimal model includes 4 of the 5 predictor variables (inter-crime distances, entry behaviours, property stolen and internal behaviours). In residential D, the optimal model includes 1 of the 5 predictor variables (inter-crime distances). Thus, the only thing all four optimal models have in common is they include inter-crime distances as a predictor variable. In addition, 3 of the 4 optimal models (A, B and C) explain a higher proportion of variance in the criterion variable than any single feature model and they fit the data better.

7.6. Logistic regression models for commercial burglary

Each of the single feature regression models constructed from the commercial district samples are presented in Table 7.7 (a, b, c and d) and will be discussed first before moving on to the optimal models. These single feature models were again constructed using direct logistic regression analysis, and include a model developed for inter-crime distances, the combined component, entry behaviours, target selection choices, property stolen, and internal behaviours.

Table 7.7 (a). Logistic regression models for commercial A

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-0.37 (0.60)	-0.09 (0.10)	0.88 (1)	0.03	0.89 (1)
Internal	-0.93 (0.36)	16.27 (73.32)	0.05 (1)	0.09	2.47 (1)
Entry	-0.66 (0.41)	-1.91 (2.53)	0.57 (1)	0.03	0.72 (1)
Target	-0.90 (0.36)	0.90 (1.46)	0.38 (1)	0.01	0.37 (1)
Property	-0.88 (0.36)	0.53 (1.42)	0.14 (1)	0.01	0.14 (1)
Combined	-0.84 (0.44)	-0.08 (3.43)	0.00 (1)	0.00	0.00 (1)
Optimal	--	--	--	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 7.7 (b). Logistic regression models for commercial B

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-2.19 (0.18)	-0.17 (0.04)	15.56 (1)***	0.08	32.74 (1)***
Target	-3.14 (0.15)	1.18 (0.31)	14.34 (1)***	0.03	11.91 (1)***
Combined	-3.38 (0.19)	2.55 (0.71)	12.77 (1)***	0.03	11.34 (1)***
Property	-3.08 (0.14)	1.45 (0.42)	11.95 (1)***	0.02	9.31 (1)**
Internal	-3.04 (0.14)	0.85 (0.46)	3.40 (1)	0.01	2.91 (1)
Entry	-3.13 (0.18)	0.76 (0.51)	2.19 (1)	0.01	2.06 (1)
Optimal	-2.39 (0.21)	--	--	0.11	48.33 (3)***
Distance	--	-0.16 (0.04)	14.01(1)***	--	--
Property	--	1.40(0.44)	10.28(1)***	--	--
Target	--	0.87(0.33)	7.18(1)**	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 7.7 (c). Logistic regression models for commercial C

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	-1.53 (0.14)	-0.91 (0.09)	104.76 (1)***	0.23	245.01 (1)***
Target	-3.86 (0.11)	1.79 (0.20)	76.76 (1)***	0.06	58.89 (1)***
Combined	-4.17 (0.14)	4.12 (0.57)	51.31 (1)***	0.04	43.62 (1)***
Entry	-3.86 (0.12)	1.35 (0.34)	15.68 (1)***	0.01	13.78 (1)***
Property	-3.66 (0.10)	0.97 (0.35)	7.74 (1)**	0.01	6.39 (1)**
Internal	-3.64 (0.10)	0.81 (0.39)	4.23 (1)	0.00	3.63 (1)
Optimal	-2.11 (0.19)	--	--	0.25	273.82 (1)***
Distance	--	-0.81(0.09)	89.61(1)***	--	--
Target	--	0.87(0.22)	15.26(1)***	--	--
Entry	--	1.06(0.36)	8.71(1)**	--	--
Property	--	0.93(0.38)	5.96(1)**	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table 7.7 (d). Logistic regression models for commercial D

Model	Constant (SE)	Logit (SE)	Wald (df)	R^2	X^2 (df)
Distance	0.40 (0.25)	-0.85 (0.11)	61.88 (1)***	0.51	195.33(1)***
Combined	-3.38 (0.21)	3.66 (0.70)	27.64 (1)***	0.07	24.81(1)***
Property	-2.91 (0.15)	1.79 (0.35)	26.15 (1)***	0.06	20.96(1)***
Target	-2.89 (0.15)	1.46 (0.33)	19.95 (1)***	0.05	16.39(1)***
Entry	-3.04 (0.21)	1.52 (0.56)	7.30 (1)**	0.02	6.69(1)**
Internal	-2.80 (0.15)	0.81 (0.37)	4.84 (1)*	0.01	4.26(1)*
Optimal	0.21 (0.26)	--	--	0.53	204.16 (2)***
Distance	--	-0.86 (0.11)	61.22 (1)***	--	--
Property	--	1.41 (0.47)	8.95 (1)**	--	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

7.6.1. The regression coefficients

Consistent with the findings presented in Table 7.4, the signs of the logit coefficients in the single feature models indicate that unlinked crime pairs tend to be characterised by larger inter-crime distances and lower levels of across-crime similarity than linked crime pairs. There are only two cases where this is not true, for the similarity scores associated with entry behaviours and the combined component in commercial A. However, the standard errors associated with the regression coefficients in these models suggest there is a general problem, which is probably due to the relatively small sample size in commercial A.

As before, to determine what these logit coefficients mean in more practical terms, each of them was multiplied by 0.10 and exponentiated. The results from this analysis are present in Table 7.8 where it can be seen that an increase of 0.10 units for inter-crime distances would cause a decrease in the odds of two crimes being linked. In contrast, an increase of 0.10 units for all other similarity scores would cause an increase in the odds of two crimes being linked, with the exception of entry behaviours and the combined component in commercial A.

Table 7.8. The odds of crimes being linked as a function of similarity

	District			
Variables	A	B	C	D
Distance	0.99	0.98	0.91	0.92
Combined	0.99[a]	1.29	1.51	1.44
Entry	0.83[a]	1.08	1.14	1.16
Property	1.05	1.16	1.10	1.20
Target	1.09	1.13	1.20	1.16
Internal	5.09	1.09	1.08	1.08

[a] Contrary to prediction

The impact of changes in the predictor variables was also assessed by examining changes in the probability that two crimes are linked. Unsurprisingly, the general effect of increasing across-crime similarity scores was an increase in the probability that two crimes are linked.

7.6.2. Predictive accuracy and goodness-of-fit

Also consistent with the findings presented in Table 7.4, many of the regression models were found to have high levels of predictive accuracy and fit. In addition, the same consistencies seen when examining residential burglaries emerge. In other words, the regression models including inter-crimes distances are always characterised by the highest levels of predictive accuracy and fit, while the regression models including internal behaviours are always characterised by the lowest levels of predictive accuracy and fit. Again, no other consistencies emerge in relation to any of the other behavioural domains.

7.6.3. The multiple feature models

In addition to constructing single feature logistic regression models, optimal regression models were also constructed (the combined component was again not included in these analyses). Summaries of the optimal logistic regression models are also presented in Table 7.7 (a, b, c and d). As was the case with residential burglary, what can be seen from these tables is that the variables included in the optimal models are dependent on the police district considered.

In commercial A, an optimal model could not be constructed due to the small sample size. In commercial B, the optimal model includes 3 of the 5 predictor variables (inter-crime distances, property stolen and target selection choices). In commercial C, the optimal model includes 4 of the 5 predictor variables (inter-crime distances, target selection choices, entry behaviours and property stolen). In commercial D, the optimal model includes 2 of the 5 predictor variables (inter-crime distances and property stolen). Thus, all 3 optimal models have in common is they include inter-crime distances. In addition, all of the optimal models explain a higher proportion of the variance in the criterion variable than any single feature model and they fit the data better.

7.7. ROC analysis

In the previous section, logistic regression models constructed from data in the experimental samples indicated that many burglary behaviours exhibited by offenders in Merseyside could be used to predict whether crime pairs are unlinked or linked. However, the accuracy of the models varied depending on the behavioural domain and police district considered. As in the previous chapters, ROC analysis was carried out on these regression models in order obtain independent measures of consistency and discrimination, as well as some indication of model validity. To carry out this analysis, the regression models presented in Tables 7.5 and 7.7 were used to calculate estimated probabilities for every possible crime pair in the residential and commercial test samples. These probabilities were then used to construct separate ROC graphs for each behavioural domain, including the optimal models.

7.8. Single feature ROC graphs for residential burglary

The results from the single feature ROC graphs for residential burglary, along with their AUCs (and *p*-values), standard errors, and 95% confidence intervals, are summarised in Table 7.9 (see Appendix J for the actual ROC graphs). These results correspond to the single feature regression models presented in Table 7.5 once they had been applied to each and every crime pair in the test samples.

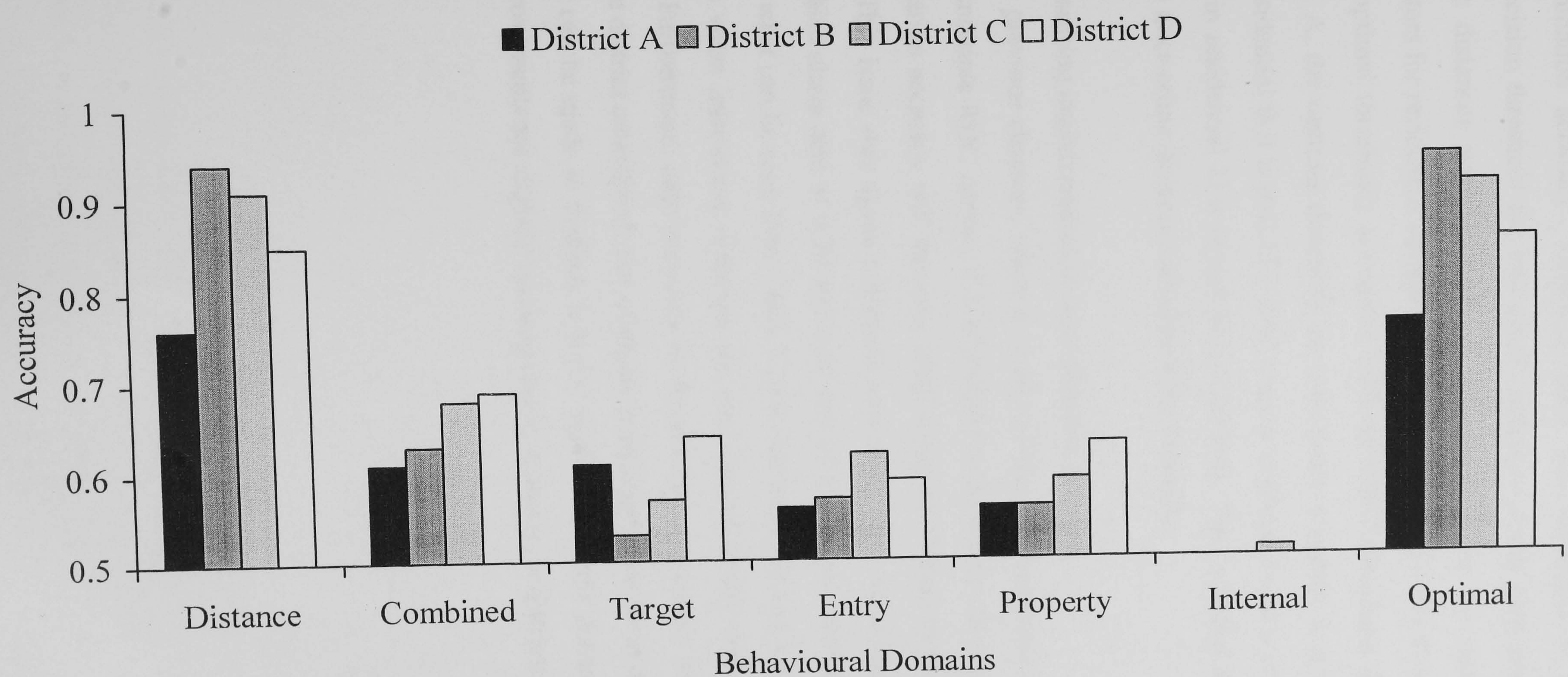
7.8.1. The AUC as a measure of spatial and behavioural consistency

Consistent with analysis of data in the experimental samples, the majority of ROC curves indicate that residential burglary behaviours in Merseyside are expressed in a consistent fashion, though not all are consistent beyond what would be expected by chance. The AUCs associated with these ROC curves have been summarised in graph form and are presented as Figure 7.2. This graph confirms that the level of consistency observed is domain and district specific. As before, two consistencies emerge from the ROC analyses and are made clear in Figure 7.2. First, inter-crime distances are always the most consistent aspect of residential burglary behaviour across the Merseyside districts. Second, internal behaviours are always the least consistent aspect of residential burglary behaviour across the Merseyside districts.

Table 7.9. Summary of the Merseyside residential ROC graphs

Variable	District	AUC	SE	95% CI
Distance	A	0.76 ($p \leq 0.001$)	0.04	0.68-0.84
	B	0.94 ($p \leq 0.001$)	0.01	0.92-0.97
	C	0.91 ($p \leq 0.001$)	0.01	0.88-0.93
	D	0.85 ($p \leq 0.001$)	0.03	0.79-0.91
Combined	A	0.61 ($p \leq 0.05$)	0.05	0.57-0.70
	B	0.63 ($p \leq 0.01$)	0.04	0.55-0.71
	C	0.68 ($p \leq 0.001$)	0.02	0.63-0.73
	D	0.69 ($p \leq 0.01$)	0.06	0.58-0.81
Target	A	0.61 ($p \leq 0.05$)	0.05	0.52-0.71
	B	0.53 ($p > 0.05$)	0.04	0.45-0.61
	C	0.57 ($p \leq 0.01$)	0.03	0.52-0.62
	D	0.64 ($p \leq 0.05$)	0.06	0.52-0.75
Property	A	0.56 ($p > 0.05$)	0.05	0.45-0.66
	B	0.56 ($p > 0.05$)	0.05	0.47-0.65
	C	0.59 ($p \leq 0.001$)	0.03	0.54-0.64
	D	0.63 ($p \leq 0.05$)	0.07	0.50-0.76
Entry	A	0.56 ($p > 0.05$)	0.05	0.47-0.66
	B	0.57 ($p > 0.05$)	0.05	0.48-0.66
	C	0.62 ($p \leq 0.001$)	0.03	0.57-0.67
	D	0.59 ($p > 0.05$)	0.07	0.45-0.73
Internal	A	0.50 ($p > 0.05$)	0.05	0.40-0.59
	B	0.47 ($p > 0.05$)	0.04	0.38-0.56
	C	0.51 ($p > 0.05$)	0.03	0.46-0.56
	D	0.48 ($p > 0.05$)	0.07	0.36-0.61
Optimal	A	0.76 ($p \leq 0.001$)	0.04	0.68-0.84
	B	0.94 ($p \leq 0.001$)	0.01	0.92-0.96
	C	0.91 ($p \leq 0.001$)	0.01	0.89-0.94
	D	0.85 ($p \leq 0.001$)	0.03	0.79-0.91

Figure 7.2. AUCs for Merseyside residential burglary data



7.8.2. Operating points as a measure of discrimination

Decision thresholds in this case can once again be seen to have a serious impact on discrimination accuracy. Youden's index was calculated to identify an optimal decision threshold for inter-crime distances across each police district. Inter-crime distances were focused on because they are the most effective discriminators for residential burglaries in Merseyside (see Table J9 in Appendix J for the optimal thresholds associated with all other behavioural domains). In residential A, the optimal threshold for inter-crime distances is $p \geq 0.12$ (≤ 1.90 km), in residential B it is $p \geq 0.15$ (≤ 2.60 km), in residential C it is $p \geq 0.04$ (≤ 2.10 km), and in residential D it is $p \geq 0.24$ (≤ 2.20 km). Thus, optimal thresholds in relation to inter-crime distances are also district specific.

7.8.3. Measuring improvements in discrimination accuracy

As in the previous chapters, when a variety of burglary behaviours are used to generate multiple ROC curves, it is of interest to determine the extent to which discrimination accuracy will improve when using one behavioural domain over another. This issue was again examined here using inter-crime distances as the ideal, a false alarm rate of 0.20 and a hit rate of 0.80. Consistent with previous analysis, what can be seen from Table 7.10 is that more hits and less false alarms are made when inter-crime distances are used instead of any other behavioural domain. Furthermore, improvements in discrimination accuracy are specific to the police district considered. For example, huge improvements in discrimination accuracy can be made in districts B and C when inter-crime distances are used. The improvements are slightly less impressive, however, in districts A and D.

Table 7.10. Improvements in discrimination accuracy

	Accuracy of inter-crime distances at:	
	pFA = 0.20	pH = 0.80
District A:		
Combined	+28 hits	-22 false alarms
Target	+29 hits	-15 false alarms
Entry	+41 hits	-18 false alarms
Property	+27 hits	-28 false alarms
Internal	+41 hits	-30 false alarms
District B:		
Combined	+65 hits	-40 false alarms
Target	+72 hits	-66 false alarms
Entry	+60 hits	-65 false alarms
Property	+66 hits	-66 false alarms
Internal	+65 hits	-73 false alarms
District C:		
Combined	+41 hits	-43 false alarms
Target	+59 hits	-60 false alarms
Entry	+56 hits	-59 false alarms
Property	+50 hits	-60 false alarms
Internal	+64 hits	-68 false alarms
District D:		
Combined	+8 hits	-24 false alarms
Target	+23 hits	-39 false alarms
Entry	+15 hits	-52 false alarms
Property	+16 hits	-45 false alarms
Internal	+40 hits	-57 false alarms

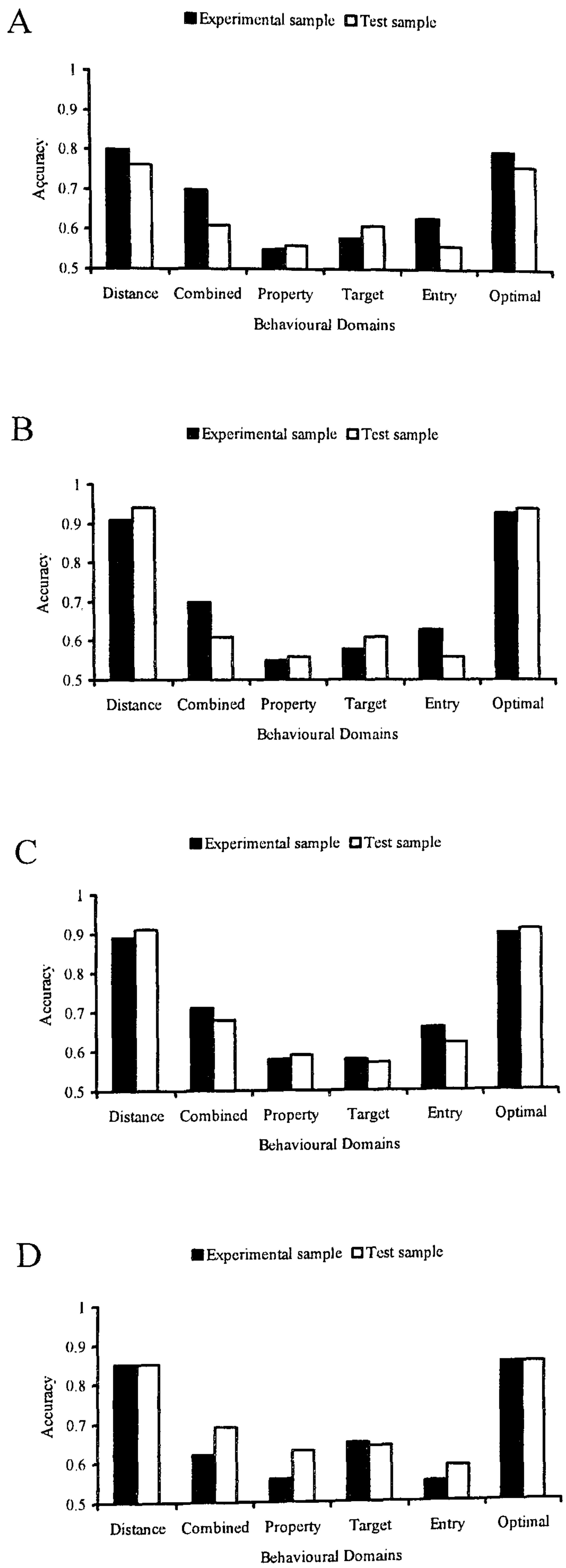
7.9. Multiple feature ROC graphs for residential burglary

In order to get some indication of how valid the optimal regression models are for residential burglaries in Merseyside, they were also used to construct ROC graphs. The results from these ROC graphs are also presented in Table 7.9 and Figure 7.2. While the optimal ROC curves are significantly higher than many single feature ROC curves, none are significantly higher than the single feature ROC curves for inter-crime distances ($p>0.05$).

7.10. Validating the empirical ROC curves for residential burglary

When the logistic regression models developed using data from the residential experimental samples are applied to data from the residential test samples, the ROC curves that result from the most accurate regression models have AUCs that are significantly greater than chance. This suggests that these logistic regression models do generalise to crimes beyond those used to construct the original models. The validity of the logistic regression models was tested more directly by constructing ROC curves using data from the experimental samples, and comparing their corresponding AUCs to the ROC results in Table 7.9. This is done in Figure 7.3 where it can be seen that the regression models perform equally well using data from the experimental and test samples. Indeed, there are many cases where the logistic regression models perform slightly better using data from the test samples.

Figure 7.3. Comparison of AUCs across the experimental and test samples



7.10.1. External discrimination trials

In addition to this validation procedure, a number of discrimination trials were carried out. Again, the goal was to determine whether the values of *p*H and *p*FA associated with points along a ROC curve correspond to the values of *p*H and *p*FA obtained across discrimination trials. In this case, only the optimal decision thresholds associated with inter-crime distances were tested since these were the most effective discriminators for residential burglaries in Merseyside. If the ROC graphs associated with these distances provide a valid representation of how accurately discriminations can be made, it should be possible to achieve similar levels of accuracy across discrimination trials when the same decision thresholds are used.

Using the same method as in previous chapters, the results are summarised in Tables 7.11 and 7.12 for random samples of various sizes, where the sample size was dependent on the number of crimes committed in each police district (see Appendix J for the full results from each discrimination trial). As can be seen by the results presented in Table 7.11, the average hit and false alarm rates observed across the small and large discrimination trials generally correspond with the predicted values, though the values for *p*H are not quite as close.

Table 7.11. Predicted versus observed values of *p*H and *p*FA

			Small sample observations		Large sample observations	
District	Target *p*H	Target *p*FA	*p*H	*p*FA	*p*H	*p*FA
A	0.70	0.30	0.78	0.30	0.75	0.30
B	0.86	0.15	0.81	0.15	0.78	0.14
C	0.84	0.16	0.79	0.16	0.80	0.17
D	0.75	0.25	0.80	0.26	0.78	0.27

Table 7.12. Validation trials for Merseyside residential burglary data

District	Threshold (distance)	Sample size	Average *p*H (freq.)	Average *p*M (freq.)	Average *p*CR (freq.)	Average *p*FA (freq.)
A	$p \geq 0.12$ (≤ 1.90 km)	100	0.78 (7.80)	0.22 (2.40)	0.70 (62.40)	0.30 (27.40)
A	$p \geq 0.12$ (≤ 1.90 km)	500	0.75 (37.80)	0.25 (12.60)	0.65 (293.00)	0.35 (156.60)
B	$p \geq 0.15$ (≤ 2.60 km)	500	0.81 (29.60)	0.19 (7.40)	0.85 (395.40)	0.15 (67.60)
B	$p \geq 0.15$ (≤ 2.60 km)	1000	0.78 (59.60)	0.22 (16.40)	0.86 (791.40)	0.14 (132.60)
C	$p \geq 0.04$ (≤ 2.10 km)	1000	0.79 (18.60)	0.21 (5.00)	0.84 (819.00)	0.16 (157.40)
C	$p \geq 0.04$ (≤ 2.10 km)	5000	0.80 (114.40)	0.20 (28.80)	0.83 (4040.60)	0.17 (816.20)
D	$p \geq 0.24$ (≤ 2.20 km)	100	0.80 (15.00)	0.20 (3.60)	0.74 (60.20)	0.26 (21.20)
D	$p \geq 0.24$ (≤ 2.20 km)	200	0.78 (26.60)	0.22 (7.40)	0.73 (121.20)	0.27 (44.80)

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

7.11. Single feature ROC graphs for commercial burglary

The results from the single feature ROC graphs for commercial burglary, along with their AUCs (and *p*-values), standard errors, and 95% confidence intervals, are summarised in Table 7.13 (see Appendix J for the actual ROC graphs). These results correspond to the single feature regression models presented in Table 7.7 once they had been applied to each and every crime pair in the test samples.

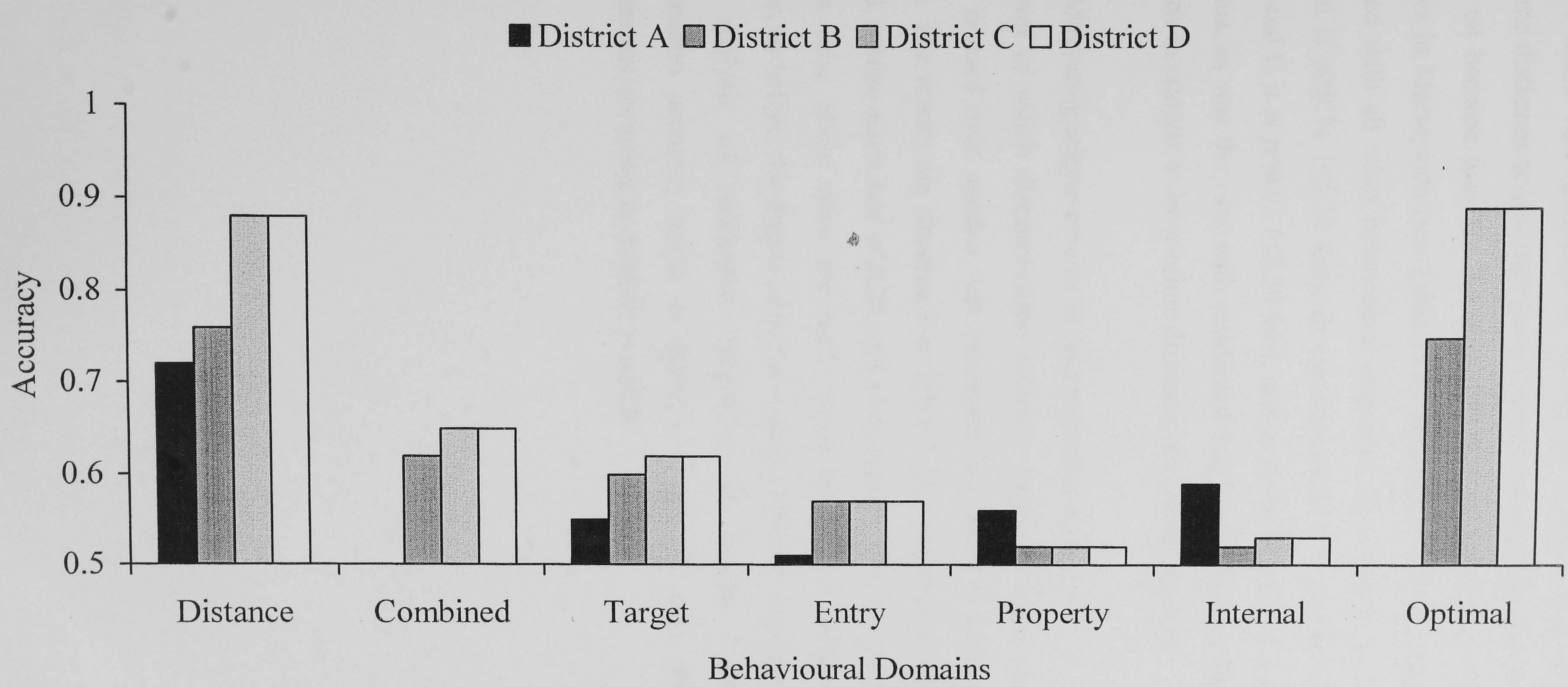
7.11.1. The AUC as a measure of spatial and behavioural consistency

Consistent with analysis of data in the commercial experimental samples, and with the earlier descriptive analysis, the majority of ROC curves indicate that commercial burglary behaviours in Merseyside are expressed in a consistent fashion, though not all are consistent beyond what would be expected by chance. The AUCs associated with these ROC curves are summarised in Figure 7.4. This graph confirms that the level of consistency observed is largely domain and district specific. As with residential burglaries in Merseyside, inter-crime distances are always the most consistent feature across each of the police districts, and property stolen and internal behaviours the least consistent.

Table 7.13. Summary of Merseyside commercial ROC graphs

Variable	District	AUC	SE	95% CI
Distance	A	0.72 ($p \leq 0.001$)	0.08	0.55-0.88
	B	0.76 ($p \leq 0.001$)	0.04	0.69-0.83
	C	0.88 ($p \leq 0.001$)	0.01	0.85-0.91
	D	0.88 ($p \leq 0.001$)	0.01	0.85-0.91
Combined	A	0.39 ($p > 0.05$)	0.11	0.19-0.60
	B	0.62 ($p \leq 0.001$)	0.04	0.54-0.70
	C	0.65 ($p \leq 0.001$)	0.02	0.60-0.70
	D	0.65 ($p \leq 0.001$)	0.02	0.60-0.70
Target	A	0.55 ($p > 0.05$)	0.11	0.33-0.71
	B	0.60 ($p \leq 0.01$)	0.04	0.52-0.68
	C	0.62 ($p \leq 0.001$)	0.03	0.56-0.67
	D	0.62 ($p \leq 0.01$)	0.03	0.56-0.67
Property	A	0.56 ($p > 0.05$)	0.11	0.35-0.77
	B	0.52 ($p > 0.05$)	0.06	0.45-0.60
	C	0.52 ($p > 0.05$)	0.03	0.47-0.58
	D	0.52 ($p \leq 0.001$)	0.03	0.47-0.58
Entry	A	0.51 ($p > 0.05$)	0.11	0.30-0.71
	B	0.57 ($p > 0.05$)	0.04	0.49-0.64
	C	0.57 ($p \leq 0.01$)	0.03	0.52-0.62
	D	0.57 ($p \leq 0.01$)	0.03	0.52-0.62
Internal	A	0.59 ($p > 0.05$)	0.11	0.38-0.81
	B	0.52 ($p > 0.05$)	0.04	0.45-0.60
	C	0.53 ($p > 0.05$)	0.03	0.47-0.58
	D	0.53 ($p > 0.05$)	0.03	0.47-0.58
Optimal	A	0.50 ($p > 0.05$)	0.11	0.30-0.71
	B	0.75 ($p \leq 0.001$)	0.04	0.68-0.82
	C	0.89 ($p \leq 0.001$)	0.01	0.86-0.91
	D	0.89 ($p \leq 0.001$)	0.01	0.86-0.91

Figure 7.4. AUCs for Merseyside commercial burglary data



7.11.2. Operating points as a measure of discrimination

The decision thresholds again have a serious impact on discrimination accuracy. Youden's index was calculated to identify an optimal decision threshold for inter-crime distances across each police district. Inter-crime distances were again focused on because they are the most effective discriminators for commercial burglaries in Merseyside (see Table J9 in Appendix J for the optimal thresholds associated with all other behavioural domains). In commercial A, the optimal threshold is $p \geq 0.30$ (≤ 5.20 km), in commercial B it is $p \geq 0.01$ (≤ 3.00 km), in commercial C it is $p \geq 0.01$ (≤ 2.30 km), and in commercial D it is $p \geq 0.17$ (≤ 2.30 km). Thus, as was the case with residential burglary in Merseyside, the optimal thresholds in relation to inter-crime distances are district specific.

7.11.3. Measuring improvements in discrimination accuracy

The extent to which discrimination accuracy improves when one behavioural domain is used over another was examined in relation to each of the district samples. The results are illustrated in Table 7.14 using inter-crime distances as the ideal, a false alarm rate of 0.20, and a hit rate of 0.80. As expected, more hits and less false alarm rates are made when inter-crime distances are used. However, as before, the degree of improvement is district dependent. In contrast to the analysis of residential burglaries, the largest improvements in discrimination accuracy occur in districts C and D, with slightly smaller improvements occurring in districts A and B.

Table 7.14. Improvements in discrimination accuracy

	Accuracy of inter-crime distances at:	
	*p*FA = 0.20	*p*H = 0.80
District A:		
Combined	+23 hits	-58 false alarms
Target	+10 hits	-42 false alarms
Entry	+16 hits	-45 false alarms
Property	+7 hits	-40 false alarms
Internal	+1 hits	-38 false alarms
District B:		
Combined	+30 hits	-17 false alarms
Target	+34 hits	-17 false alarms
Entry	+42 hits	-18 false alarms
Property	+47 hits	-21 false alarms
Internal	+47 hits	-21 false alarms
District C:		
Combined	+55 hits	-39 false alarms
Target	+42 hits	-55 false alarms
Entry	+56 hits	-55 false alarms
Property	+58 hits	-61 false alarms
Internal	+58 hits	-61 false alarms
District D:		
Combined	+48 hits	-79 false alarms
Target	+55 hits	-78 false alarms
Entry	+58 hits	-78 false alarms
Property	+51 hits	-78 false alarms
Internal	+65 hits	-82 false alarms

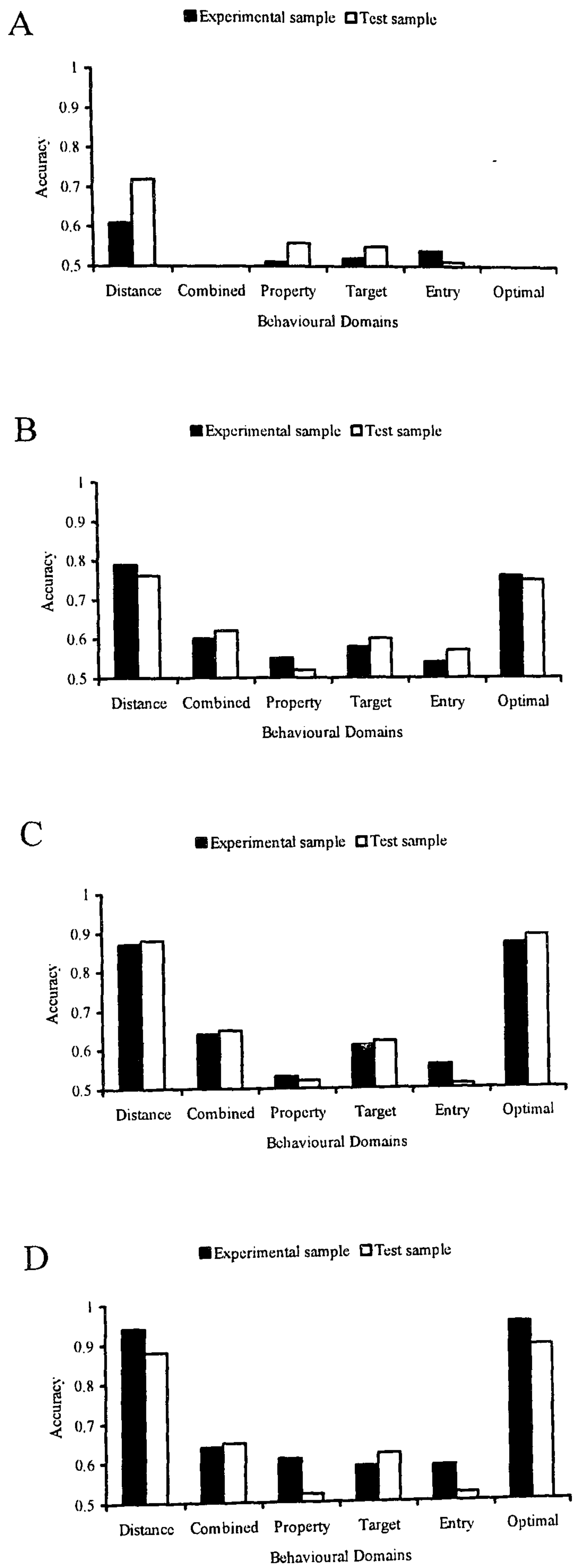
7.12. Multiple feature ROC graphs for commercial burglary

In order to get some indication of how valid the optimal regression models were for commercial burglaries in Merseyside, they were also used to construct ROC graphs. The results from these ROC graphs are also presented in Table 7.13 and Figure 7.4. Consistent with the analysis of Merseyside residential burglaries, none of the optimal ROC curves are significantly higher than the single feature ROC curves for inter-crime distances ($p>0.05$).

7.13. Validating the empirical ROC curves for commercial burglary

When the logistic regression models developed from the commercial experimental samples are applied to data from the commercial test samples, the ROC curves that result from the most accurate regression models have AUCs that are significantly greater than chance. This suggests that these logistic regression models do generalise to crimes beyond those used to construct the original models. The validity of the logistic regression models was tested more directly by constructing ROC curves using data from the experimental samples, and comparing their corresponding AUCs to the ROC results in Table 7.13. This is done in Figure 7.5 where it can be seen that the regression models perform equally well using data from the experimental and test samples.

Figure 7.5. Comparison of AUCs across the experimental and test samples



7.13.1. External discrimination trials

In addition to this validation procedure, a number of discrimination trials were also carried out. Again, only the ROC curves associated with inter-crime distances were tested, since these were the most effective discriminators for commercial burglaries in Merseyside. If the ROC graphs for these distances provide a valid representation of how accurately discriminations can be made, it should be possible to achieve similar levels of accuracy across random discrimination trials when the same decision thresholds are used.

The results of these trials are summarised in Tables 7.15 and 7.16 for random samples of various sizes, where the sample size was dependent on the number of crimes committed in each police district (see Appendix J for the full results from each discrimination trial). As can be seen from the results presented in Table 7.15, the average hit and false alarm rates observed across the small and large discrimination trials generally correspond with the predicted values, though there are some instances where the observed values are not so close.

Table 7.15. Predicted versus observed values of *p*H and *p*FA

			Small sample observations		Large sample observations	
District	Target *p*H	Target *p*FA	*p*H	*p*FA	*p*H	*p*FA
A	0.63	0.37	0.58	0.45	0.52	0.42
B	0.75	0.25	0.78	0.37	0.77	0.38
C	0.81	0.19	0.88	0.28	0.88	0.29
D	0.90	0.10	0.82	0.09	0.83	0.09

Table 7.16. Validation trials for Merseyside commercial burglary data

District	Threshold (distance)	Sample size	Average pH (freq.)	Average pM (freq.)	Average pCR (freq.)	Average pFA (freq.)
A	$p \geq 0.30$ (≤ 5.20 km)	10	0.58 (1.80)	0.42 (1.60)	0.55 (3.80)	0.45 (2.80)
A	$p \geq 0.30$ (≤ 5.20 km)	50	0.52 (7.20)	0.48 (6.60)	0.58 (20.80)	0.42 (15.40)
B	$p \geq 0.01$ (≤ 3.00 km)	1000	0.78 (35.20)	0.22 (9.80)	0.63 (603.60)	0.37 (351.40)
B	$p \geq 0.01$ (≤ 3.00 km)	2000	0.77 (75.60)	0.23 (22.60)	0.62 (1186.40)	0.38 (715.40)
C	$p \geq 0.01$ (≤ 2.30 km)	1000	0.88 (22.00)	0.12 (3.00)	0.72 (695.20)	0.28 (279.80)
C	$p \geq 0.01$ (≤ 2.30 km)	5000	0.88 (119.40)	0.12 (15.40)	0.71 (3439.00)	0.29 (1426.20)
D	$p \geq 0.17$ (≤ 2.30 km)	500	0.82 (28.40)	0.18 (6.40)	0.91 (421.20)	0.09 (44.00)
D	$p \geq 0.17$ (≤ 2.30 km)	1000	0.83 (55.00)	0.17 (11.00)	0.91 (848.40)	0.09 (85.60)

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

7.14. Chapter summary

In this chapter, the behaviour of residential and commercial burglars committing crimes in Merseyside was explored. The data collected from Merseyside extended the analysis of burglary behaviour presented in the previous chapters in an important way. Unlike the data collected from London, Dorset and Oldham, the Merseyside burglary data allowed different groups of serial burglars committing crimes in the same force area to be compared with one another based on the police district in which they committed their crimes.

A descriptive analysis of the across-crime similarity scores provided partial support for the hypothesis that offenders committing burglaries in the Merseyside districts exhibit their offence behaviours in a consistent fashion. In the case of residential burglary, highly significant differences were generally found in each of the four police districts for every behavioural domain except internal behaviours. In the case of commercial burglary, highly significant differences were generally found for every behavioural domain, though not usually in the case of Merseyside A.

Logistic regression analysis was run on the similarity scores associated with each behavioural domain as a way of calculating predicted probabilities that crime pairs were linked. These probabilities, in turn, formed the basis of ROC analysis. Consistent with the descriptive analysis, ROC analysis indicated that inter-crime distances are the most consistent aspect of residential and commercial burglary across every district in Merseyside. In line with this, relatively high levels of discrimination accuracy could be achieved in every district when using these distances. Similar results emerged across discrimination trials, suggesting that the ROC procedure has an adequate level of predictive accuracy in Merseyside.

The results presented in this chapter are consistent with the results from the previous chapter in a number of ways. First, the results confirm that residential and commercial burglars are spatially consistent with respect to their inter-crime distances, though in Merseyside the level of consistency appear to be slightly higher for residential burglaries. Second, the results confirm that the use of inter-crime distances leads to higher levels of discrimination accuracy compared to any other crime scene behaviour, though in Merseyside the optimal decision thresholds have to be slightly more lenient for commercial burglaries. Third, the results confirm that these high levels of discrimination accuracy can be achieved even when the burglaries are sampled at a district level, though in Merseyside the levels of discrimination accuracy is district dependent.

The results presented in this chapter also extend the findings presented in previous chapters in an important way. It is clear from the Merseyside results that while inter-crime distances are the best discriminators across all of the Merseyside districts, in order to achieve the highest possible levels of discrimination accuracy within each district, different decision thresholds have to be adopted. From a practical perspective this is extremely important. It suggests that a general strategy for linking serial burglaries, developed at the regional or even force-wide level, will not be the most effective approach to take. This stands in direct contrast to proposals put forward for dealing with other types of crime such as serial rape and murder. In these cases, state-wide and nation-wide discrimination strategies are still very much the norm (e.g., Grubin *et al.*, 2001; Keppel & Weis, 1993).

CHAPTER 8
DISCUSSION AND CONCLUSIONS

8.1. Introduction

The present research had two primary objectives. The first objective was to examine the conditions under which serial burglars display consistent patterns of behaviour across the crimes they commit. The second objective was to examine the extent to which these consistencies, if they did in fact exist, could be used to discriminate between crimes committed by different offenders. The only way these two objectives could be met, it was argued, was to develop appropriate operational definitions of each behavioural process. Only once this was done could an effective analytical procedure be developed to accurately measure the extent to which consistency exists and the degree of discrimination that was possible as a result.

8.2. Defining consistency and discrimination in the criminal context

In the present research, consistency and discrimination are assumed to be the result of various other behavioural processes. Behavioural consistency is viewed as a product of how similar the behaviours are that serial burglars exhibit across their own crimes and how distinct these behaviours are when compared to behaviours exhibited by other offenders committing similar sorts of crimes. As a result of defining behavioural consistency in this way, the possibility of accurately discriminating between crimes committed by different offenders can be viewed as a product of how consistent serial burglars are. Essentially, the higher the level of consistency exhibited by burglars within a given sample, the higher the level of discrimination that should be possible as a result.

In line with these definitions, it was also claimed that each of the various behavioural processes could be thought of in terms of underlying probability distributions of across-crime similarity scores. Thinking about the processes in this way made it possible to see exactly how each process would emerge in observable burglary behaviours and to see how each process relates to and depends on the other. This, in turn, led to interesting ideas about how behavioural consistency and discrimination could effectively be measured, so long as a suitable analytical procedure could be developed.

8.3. Measuring consistency and discrimination in the criminal context

In the present research, it has been argued that behavioural consistency can be measured by quantifying the degree of overlap between intra-offender and inter-offender similarity scores. Additionally, it has been argued that while overall levels of behavioural discrimination can be measured in the same way, threshold-specific levels of discrimination accuracy can be measured by quantifying the proportions of different decision outcomes resulting at specific decision thresholds. The challenge, then, was to come up with an analytical procedure that could effectively measure all these things.

A procedure based on ROC analysis proved to be a suitable approach. However, it is important to stress that the only way ROC analysis can be used effectively for this purpose is to define consistency and discrimination in the way they have been here. Indeed, when consistency and discrimination are defined in this way, ROC analysis can be shown to relate directly to each of the behavioural processes. Specifically, the area under the ROC curve (AUC) reflects the degree of overlap between intra-offender and inter-offender similarity scores, thus providing a measure of consistency and overall discrimination. In addition, the points falling along a ROC curve reflect the proportions of decision outcomes resulting at each decision threshold, thus providing a measure of threshold-specific discrimination.

The use of ROC analysis in this context represents a step forward for two primary reasons. First, never before has a procedure been proposed in the criminal context to identify the conditions under which consistency and discrimination become most apparent. It is now possible to identify at a fairly precise level not only whether one set of behaviours will have more discriminatory power than another, but also how much more discriminatory these behaviours will be. Second, never before has the importance of decision thresholds been mentioned in the criminal context, nor a method proposed to determine how they should be set. It is now possible to not only decide when two crimes are similar enough to warrant being linked, but also to determine what the consequences will be if different decision thresholds are used.

8.4. Behavioural consistency in the criminal context

Having examined the levels of consistency expressed by serial burglars in the UK, a comparison can now be made between the different samples examined. This comparison will allow general patterns of consistency to be identified, which should help answer questions about how burglars in the UK commit their crimes. Before proceeding with this comparison, however, it is important to emphasise the point that the only reason an appropriate comparison can be made in the first place is because of the advantages associated with the various ROC measures used throughout this research.

To reiterate, there are a number of advantages associated with the AUC, including the fact that the AUC is a single, flexible, and general measure of consistency. The AUC provides a single measure of consistency because it corresponds to the position of an entire ROC curve, rather than being dependent on any specific threshold. The AUC provides a flexible measure of consistency because it can be calculated, regardless of what behaviour is being observed or how it is being measured. The AUC provides a general measure of consistency because it is not based on the relative frequencies or base rates of unlinked and linked crime pairs in any particular sample but on their proportions.

To compare the levels of behavioural consistency exhibited by serial burglars across the samples, these AUC measures will be focused on. Table 8.1 presents a summary of the AUCs calculated in each of the previous chapters for each of the behavioural domains. Average AUCs are also provided for each behavioural domain, and the AUCs that meet Swets' (1988) criteria for useful discriminators are also indicated. This summary table highlights a number of interesting patterns in burglary behaviour found across the UK.

Table 8.1. Summary of the AUCs

Variables	Average AUC	London	Dorset	Oldham residential	Oldham commercial	Merseyside residential				Merseyside commercial			
						A	B	C	D	A	B	C	D
Distance	0.85[a]	0.97[a]	0.89[a]	0.80[a]	0.82[a]	0.76[a]	0.94[a]	0.91[a]	0.85[a]	0.72[a]	0.76[a]	0.88[a]	0.88[a]
Combined	0.63	--	0.67	0.65	0.68	0.61	0.63	0.68	0.69	0.39	0.62	0.65	0.65
Target	0.59	--	0.59	0.60	0.59	0.61	0.53	0.57	0.64	0.55	0.60	0.62	0.62
Entry	0.58	--	0.64	0.53	0.64	0.56	0.57	0.62	0.59	0.51	0.57	0.57	0.57
Property	0.57	--	0.52	0.64	0.60	0.56	0.56	0.59	0.63	0.56	0.52	0.52	0.52
Internal	0.52	--	0.57	--	--	0.50	0.47	0.51	0.48	0.59	0.52	0.53	0.53
Optimal	0.82[a]	--	0.90[a]	0.80[a]	0.83[a]	0.76[a]	0.94[a]	0.91[a]	0.85[a]	0.50	0.75[a]	0.89[a]	0.89[a]

[a] Useful behavioural domains based on the Swets (1988) criteria; --: No data available to calculate the AUC

8.4.1. The consistency of inter-crime distances

The first obvious pattern that emerges from Table 8.1 is that inter-crime distances are the most consistent aspect of burglary behaviour across every sample examined. The average AUC for inter-crime distance calculated across all the samples is 0.85, reflecting a relatively high degree of behavioural consistency. Indeed, according to the criteria set out by Swets (1988), all the AUCs associated with inter-crime distance fall into the category of being either moderately or highly accurate, suggesting that the degree of overlap between intra-offender and inter-offender distances in every sample is relatively small.

In more practical terms, these AUC values suggest that the choice of where to commit burglaries is usually made in a consistent fashion by the burglars included in the present research. In other words, many burglars tend to stick to a similar geographic area when committing their own crimes over time and these geographic areas tend to differ across different serial burglars. This appears to be the case for residential and commercial burglaries, committed in rural or urban areas, at different levels of geographic precision. Having said this, however, it is important to point out that the AUCs associated with inter-crime distances are also highly variable across the different samples, ranging from 0.72 to 0.97.

At a theoretical level, this finding adds something to the growing body of literature, which indicates that offenders are often limited in terms of their spatial mobility. Offender spatial behaviour is typically examined in terms of how far offenders travel from home to commit their crimes (e.g., Baldwin & Bottoms, 1976; Brantingham & Brantingham, 1981; Davies, 1996; Rengert & Wasilchick, 2000). The results presented throughout this research take the understanding of criminal mobility a stage further by clearly demonstrating that burglars living in the same general area may have relatively distinct areas of criminal activity that do not overlap to a great extent.

This finding accords well with Grubin *et al.*'s (2001) recent study of rapists. Grubin and his colleagues showed that using information about where rapists commit their crimes significantly enhances the accuracy with which offences committed by different offenders could be distinguished from one another.

However, Grubin's research suffers from the potential artefact that the rape series examined were drawn from all over the UK. Distinguishing such series by the locality in which the offences occur is therefore not as stringent a test as that undertaken in the present research. Throughout the present research, burglaries were always sampled from much smaller areas of the country. Despite this more focused sampling, inter-crime distances in serial burglary were still found to be remarkably consistent.

(a) Practical explanations

There are at least three general explanations for why inter-crime distances are more consistent than any other aspect of burglary behaviour. On the purely practical side, the fact that inter-crime distances are found to be more consistent probably tells us something about how police data is collected and recorded. The location of crime sites in each of the samples can be recorded in a very reliable and accurate fashion, which may allow consistent patterns of spatial behaviour to clearly emerge. A similar level of reliability and accuracy is not likely to be associated with the other burglary behaviours included in this research and, therefore, a degree of consistency in relation to these behaviours is probably lost amidst data error (Grubin *et al.*, 2001).

Also on a practical note, the various forms of potential bias discussed in Chapter 3 cannot be ignored as possible explanations for why inter-crime distances are so consistent. For example, it is possible that if a burglar commits some of his crimes close together but a few more some distance away, that the crimes in close proximity will have a higher chance of being linked and solved. Consequently, by only considering solved crimes in the present research the degree of consistency associated with inter-crime distances may artificially increase. It may appear that the crime site locations chosen by each offender are highly similar and fairly distinct, but this may be because the crimes that do not fit this pattern are inadvertently ignored.

(b) Psychometric explanations

Another possible explanation for this finding is that the rates of consistency observed in the present research are related to levels of across-crime variance

(Wiggins, 1973). Certainly, consistency correlations in the non-criminal context are constrained by the variances of the variables they correlate. In their study of behavioural consistency, for example, Funder and Colvin (1991) found that behavioural items with larger across-subject variances yielded higher consistency correlations across different situations. Thus, the fact that the average across-crime variances associated with each behavioural domain in the present research seems to be positively correlated with observed levels of consistency suggests this may be a plausible explanation for the differences in consistency levels.

In large part, these differences in variance are probably due to two things. First, the range of possible actions that a burglar can exhibit in relation to their spatial behaviour is much greater than the range of possible actions that a burglar can exhibit in relation to their other crime scene behaviours. For example, a burglar can travel to many different geographic areas in order to commit his crimes but the range of entry behaviours that he can exhibit is much more limited. Second, the measures used to quantify across-crime similarity in relation to inter-crime distances allows this vast range of spatial possibilities to be revealed. This is not the case for the similarity scores based on Jaccard's coefficient.

(c) Psychological explanations

There are also possible psychological explanations for why inter-crime distances are exhibited in a more consistent fashion than other burglary behaviours. For example, one explanation recently proposed by Bennell and Canter (in press) is that compared to other burglary behaviours, inter-crime distances are less situation-dependent. In other words, the choice of where to commit a burglary is determined by the offender more so than any other burglary behaviour, primarily because these sorts of decisions are often made before the offence actually takes place (Wright & Decker, 1994). As a result, it is possible that crime site selection will be unaffected by the sorts of situational influences that may cause other burglary behaviours to be exhibited in a relatively inconsistent fashion across an offender's crimes.

This line of thinking is certainly consistent with studies of non-criminal consistency. For example, as expected from various theoretical accounts (e.g., McClelland, 1984; Skinner, 1966), behaviours rated as more 'operant' by a sample of individuals are typically exhibited in a more consistent fashion when compared to behaviours that are rated as more 'respondent'. This makes sense when one considers that operant behaviours are typically defined as behaviours emitted by individuals, whereas respondent behaviours are defined as behaviours that require specific, eliciting stimuli found within particular situations in order to be expressed (Funder & Colvin, 1991).

8.4.2. The consistency of other behavioural domains

A second pattern that emerges from Table 8.1 is that the other behavioural domains examined in the present research are exhibited in a less consistent fashion than inter-crime distances, though most are consistent to a degree. This is the case for residential and commercial burglaries, committed in rural or urban areas, and at different levels of geographic precision. Specifically, the average AUCs associated with these domains range from 0.52 to 0.63. According to the Swets (1988) criteria, all these AUCs fall into the category of being either non-informative or slightly accurate. Additionally, as with inter-crime distances, the AUCs associated with these domains are variable across samples and there are no obvious relationships between the consistency observed and the sample examined.

While relatively low, the levels of consistency associated with entry behaviours, target selection choices, internal behaviours and property stolen are generally in line with existing research. Although some research indicates otherwise (e.g., Green *et al.*, 1976), most research suggests that crime scene behaviours are often exhibited in a dissimilar fashion across an offender's crimes due to various internal and external factors (Davies, 1992; Douglas & Munn, 1992; Grubin *et al.*, 2001; Turvey, 2000). Existing research also suggests that even if crime scene behaviours are exhibited in a similar fashion across crimes this does not mean that they are necessarily distinct (Grubin *et al.*, 2001; Rengert & Wasilchick, 2000; Walsh, 1980; Wright & Decker, 1994). Either of these findings could account for the low levels of consistency associated with crime scene behaviours found in the present research.

The fact that the level of consistency associated with these behaviours varies across different burglary samples also supports the idea that these behaviours are more context-dependent than inter-crime distances. As briefly mentioned above, the levels of consistency reported in the present research provide preliminary evidence that the differences in consistency across the domains relate to how situation-dependent the behaviours are. Essentially, the more an offender can decide to exhibit a set of behaviours, the higher the level of consistency becomes. Based on this reasoning it should come as no surprise that internal behaviours and property stolen are associated with the lowest levels of consistency, since these behaviours depend to a large extent on what is encountered by the offender once they have gained access to a property. Likewise, it should be expected that entry behaviours and target selection choices are associated with more moderate levels of consistency and inter-crime distances with the highest levels, because this appears to be the relative order of how situationally dependent these behaviours are.

In addition to this possible explanation, the levels of consistency associated with crime scene behaviours in the present research appear to relate to the degree of reliability and accuracy with which the data can be collected and coded in the first place. Again, inter-crime distances may be consistent, in part, because the location where burglaries are committed can be recorded with a relatively high degree of reliability and accuracy. Likewise, the two least consistent aspects of burglary behaviour in the present research, internal behaviours and property stolen, are probably the two domains that can be recorded with the least amount of reliability and accuracy.

As an example, consider the potential problems that may be encountered by the police when attempting to determine the type of property an offender has stolen. Unlike inter-crime distances, or even target selection choices and entry behaviours, the collection and recording of this information in official police records may be highly unreliable and inaccurate. This could be the case because of what the police choose to record as stolen, because of what the owner chooses to say was stolen, because of what items were available to be stolen, and because it was not clear how to record what was actually stolen.

8.4.3. Patterns of consistency when domains are combined

A third pattern that emerges from Table 8.1 relates to the combined behavioural component. In general, when behavioural domains are collapsed to form the combined component, the AUCs are higher than the AUCs associated with any behavioural domain in isolation. The only instance where this is not the case is for commercial burglaries committed in Merseyside A, though this is most likely due to a low sample size. Excluding this sample, the average AUC for the combined component is 0.65, which according to Swets' (1988) criteria is slightly accurate. Thus, although burglars exhibit relatively low levels of consistency when specific facets of their crime scene behaviour are examined in isolation, their behaviour as a whole may be slightly more consistent.

It is important to point out, however, that even though combining behavioural domains results in increased levels of behavioural consistency, these are usually less than the levels associated with inter-crime distances. Perhaps somewhat surprisingly, this general pattern also holds true for optimal combinations of burglary behaviour as well. As would be expected, optimal combinations of behaviour always result in higher levels of consistency compared to any single behavioural domain, including the combined component. However, the levels of consistency associated with these optimal combinations rarely exceed those associated with inter-crime distances. In fact, excluding the commercial burglaries committed in Merseyside A because of the small sample size, the average AUC for the optimal domain is 0.85. This is equal to the average AUC for inter-crime distances.

8.4.4. Patterns of consistency across other aspects of the data

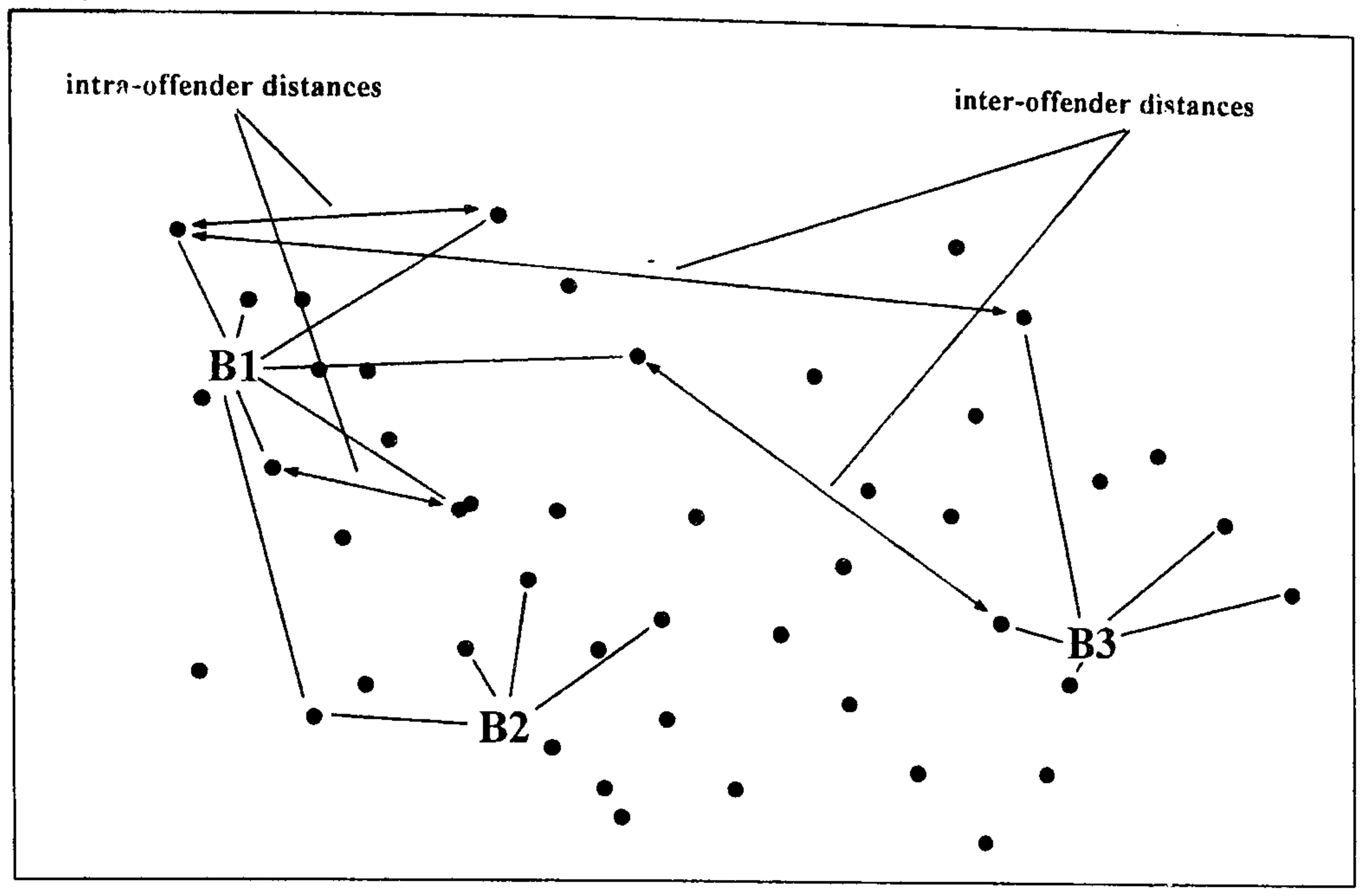
Based on the data collected for the present research, it is also possible to compare, at a very general level, how consistent burglars are across other aspects of the data. These various aspects include whether the burglaries are residential versus commercial, whether the burglaries have been committed in a rural versus urban area, and whether the burglaries have been collected at the force-wide or divisional/district level. The AUCs in Table 8.1 indicate that clear patterns of behavioural consistency rarely emerge across these various aspects, and when they do, it is usually only for inter-crime distances. As a result, only inter-crime distances will be focused on in the next 3 sub-sections.

(a) Residential versus commercial burglary

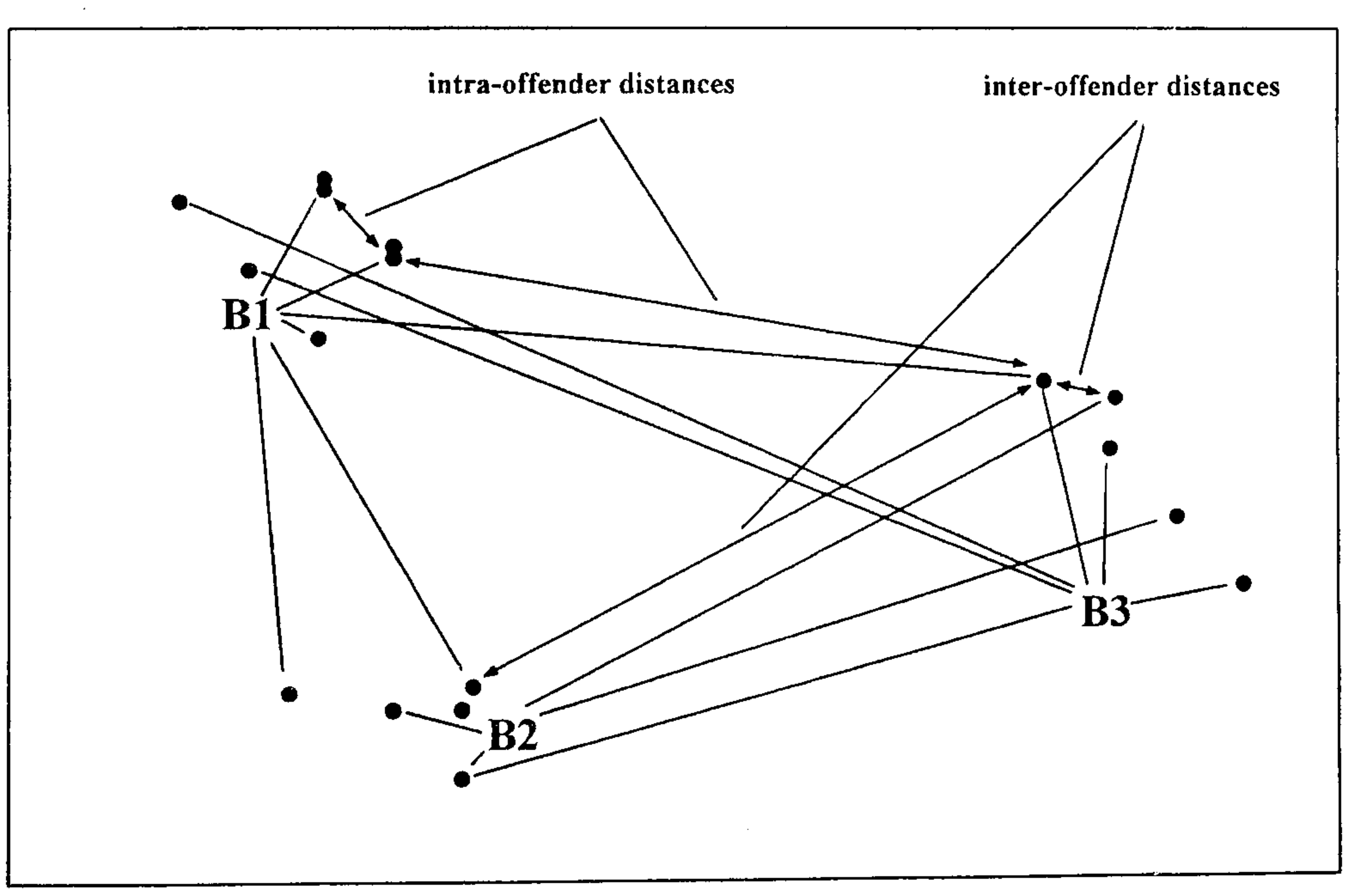
Some researchers have suggested that the spatial behaviour of serial burglars may be effected to some extent by the type of burglary being committed (e.g., Capone & Nichols, 1976; Van Koppen & Jansen, 1998). This makes sense if one assumes that there are typically more residential properties in any given geographic area compared to commercial properties, or that residential properties are more evenly distributed within these areas. In the first instance, commercial burglars would presumably have to travel to a wider variety of geographic areas than residential burglars would, if for no other reason than to reduce their chances of being detected. In the second instance, commercial burglars as a group would presumably be more likely than residential burglars would be to travel to similar geographic areas on occasion because, spatially, the opportunities to commit their crimes are somewhat more limited.

The possible impact of such a situation is illustrated in Figure 8.1. This figure contains two hypothetical maps indicating the frequency and distribution of potential targets in residential and commercial burglary. In addition, the predicted travel patterns of 3 serial burglars, B1, B2 and B3, are superimposed on each map. Consistent with what has been discussed, the frequency of potential targets in these maps is clearly greater in the case of residential burglary and the targets are more evenly distributed. The hypothesised result of such a situation is that, compared to commercial burglars, residential burglars will be more likely to target similar, yet distinct, geographic areas across the crimes they commit. Numerically, this would be reflected in shorter intra-offender distances and longer inter-offender distances for crimes committed by residential burglars, which in turn would manifest itself in higher AUCs being associated with these types of crimes.

Figure 8.1. Hypothesised behaviour of residential and commercial burglars



(a) Spatial behaviour in residential burglary



(b) Spatial behaviour in commercial burglary

To examine whether such a pattern actually does exist in the present research, the AUCs for inter-crime distances have been separated into the residential and commercial samples in Table 8.2. These AUCs suggest that a strong relationship between consistency and type of burglary does not exist in the present research, though there is a slight tendency for higher AUCs to be associated with residential burglaries. For example, in the case of Merseyside burglaries, the only district where inter-crime distances are more consistent in commercial burglary is in Merseyside D (St. Helen's and Knowsley), otherwise inter-crime distances are always more consistent in the residential samples. It must be said, however, that the opposite pattern is true for Oldham burglaries, while useful comparisons cannot be made in London and Dorset.

Table 8.2. Consistency across residential and commercial burglaries

Sample	Residential AUC	Commercial AUC
London	0.97	--
Dorset	0.89	--
Oldham	0.80	0.82
Merseyside A	0.76	0.72
Merseyside B	0.94	0.76
Merseyside C	0.91	0.88
Merseyside D	0.85	0.88

--: No data available to calculate the AUC

(b) Rural versus urban areas

Research has also suggested that there may be a relationship between the spatial behaviour of serial burglars and the areas where they commit their crimes. For example, some researchers have suggested that burglars in rural areas will travel further to commit their crimes compared to burglars who are operating in more urban areas (Van Koppen & Jansen, 1998). Presumably, this would be the case because there are less potential targets for burglary in rural areas and these targets are less evenly distributed. Indeed, based on the same argument presented above, if the frequency and distribution of potential targets do differ between rural and urban areas in this way, differences in the level of consistency associated with inter-crimes distances may also be expected.

Based on these assumptions, the expectation is that spatial consistency levels should be slightly lower in rural burglaries. As hypothesised in the case of commercial burglary, due to the lower frequency of potential targets in rural areas, rural burglars would presumably have to commit their crimes across a wider range of geographic areas to reduce their chances of being detected. Also, due to the uneven distribution of potential targets, rural burglars as a group would presumably have to travel to similar geographic areas on occasion because of the limited spatial opportunities.

Since precise measures of population density in relation to each burglary location have not been collected in the present research it is very difficult to assess whether such differences really exist between rural and urban burglaries. However, at a very general level, the relationship between population density and spatial consistency can be examined. To do this, the AUCs for inter-crime distance from Table 8.1 are presented in Table 8.3, where they have been ordered with respect to the estimated population density of the samples from which they were derived. These estimated population densities have been obtained from McGinty and Williams (2001).

Table 8.3. Consistency across rural and urban areas

Sample	Density[a]	Residential AUC	Commercial AUC
London	4611	0.97	--
Merseyside C	4053	0.91	0.88
Merseyside D	1875	0.85	0.88
Merseyside A	1875	0.76	0.72
Oldham	1551	0.80	0.82
Merseyside B	1467	0.94	0.76
Dorset	153	0.89	--

[a] The population densities for each police district are measured in persons/km^2; -- : No data available to calculate the AUC

From the results presented in Table 8.3, it can be seen that very little evidence exists for such a relationship. Apart from the residential burglaries in London being in the right position, no other burglary samples really follow the expected pattern. For example, the residential burglaries committed in Merseyside B (Sefton), which is thought to be the least densely populated police district in Merseyside, are associated with the highest AUC out of all the Merseyside districts. In addition, burglaries committed in Dorset, the area with the lowest expected population density, are associated with a much higher AUC than many of the other burglary samples.

There are a number of explanations for why a strong relationship between the area where crimes are committed and spatial consistency is not found in the present research. First, it may simply be that a strong relationship between population density and spatial consistency does not exist. Second, it may be that more specific measures of population density are required before a relationship can be found. It is certainly the case, for example, that parts of Dorset are as densely populated as parts of London, Oldham and Merseyside. Without knowing precisely what these densities are and exactly where burglaries are committed it may prove very difficult to find a relationship. Third, it may be that population density is not what spatial consistency relates to. However, drawing on other measures of density to look for a relationship, such as social densities or target densities, may prove more productive.

(c) Force-wide versus divisional/district level burglaries

Grubin *et al.* (2001) found that they were able to use information about where rapists commit their crimes in order to discriminate between rapes committed by different offenders. As pointed out earlier in this chapter, this finding is not overly surprising because the sample of rapists used in that study were drawn from all over the UK. Surely, it would be a relatively rare event for a rapist from the northern most part of England to cross paths with a rapist from the opposite end of the country. As a result, the distances between the crimes committed by these two different rapists should be able to reliably distinguish one offender from the other.

Consistent with this line of thinking, the level of geographic precision at which data was collected in the present research may be expected to relate to the degree of consistency found in relation to inter-crime distances. Specifically, burglaries collected at the force-wide level (London and Dorset) might be expected to have higher AUCs compared to burglaries collected and analysed at the divisional or district level (Oldham and Merseyside). Yet, when the level of geographic precision in the present research was increased, this was found not to be the case.

Indeed, relatively stable levels of spatial consistency were found in each burglary sample regardless of the size of area examined. In fact, the levels of consistency associated with inter-crime distances are markedly higher in certain districts of Merseyside compared to some of the samples selected from much larger geographic areas. For example, the AUC found for residential burglaries in Merseyside C (Liverpool) is higher than the AUC found for residential burglaries in Dorset, even though Dorset covers an area that is approximately 6 times as large. As a result, it can be concluded that even at a relatively local level, serial burglars show a high degree of consistency in terms of where they choose to commit their crimes.

8.5. Behavioural discrimination in the criminal context

As is the case when using the AUC as a measure of behavioural consistency, there are a number of advantages associated with using ROC-related measures to quantify behavioural discrimination. First, in addition to being a measure of consistency, the AUC also provides a measure of overall discrimination. This is because there is a direct relationship, so long as the definitions proposed in the present research are used, between the consistency with which behaviours are expressed by a sample of burglars and the level of discrimination that is possible as a result. Second, ROC analysis provides a way of measuring threshold-specific levels of discrimination by calculating the proportions of all possible decision outcomes resulting at each and every decision threshold. In turn, these calculations can be used to identify optimal decision thresholds.

Since AUCs correspond to overall measures of behavioural discrimination, all of the points just discussed in relation to behavioural consistency also hold true. Essentially, if somebody wishes to discriminate between crimes committed by different offenders as accurately as possible, the various patterns of consistency found in the present research can be of great use. At a basic level, discrimination accuracy would be at its highest when the most consistent behaviours are used. Thus, it is clear from the results in the present research that the highest level of discrimination accuracy in every burglary sample will be achieved when using inter-crime distances.

In this section, instead of concentrating on overall measures of behavioural discrimination, the focus will be on the various patterns of discrimination that emerged throughout the present research in relation to optimal decision thresholds. Optimal decision thresholds will be focused on here instead of overall discrimination levels since they are somewhat independent of the consistency patterns just discussed. To do this, the optimal decision thresholds calculated in each of the previous chapters for each of the behavioural domains are summarised in Table 8.4. This summary highlights a number of interesting patterns found across the UK.

Table 8.4. Summary of the optimal decision thresholds

Variables	London	Dorset	Oldham residential	Oldham commercial	Merseyside residential				Merseyside commercial			
					A	B	C	D	A	B	C	D
Distance	≤4.60	≤3.50	≤2.00	≤2.50	≤1.90	≤2.60	≤2.10	≤2.20	≤5.20	≤3.00	≤2.30	≤2.30
Combined	--	≥0.30	≥0.22	≥0.28	≥0.23	≥0.18	≥0.20	≥0.23	--[a]	≥0.08	≥0.05	≥0.17
Entry	--	≥0.36	≥0.20	≥0.28	≥0.20	≥0.01	≥0.06	≥0.18	--[a]	≥0.21	≥0.23	≥0.18
Target	--	≥0.35	≥0.50	≥0.48	≥0.80	≥0.70	≥0.77	≥0.40	--[a]	≥0.01	≥0.10	≥0.18
Property	--	≥0.18	≥0.18	≥0.35	≥0.02	≥0.01	≥0.12	≥0.01	≥0.02	≥0.01	≥0.01	≥0.01
Internal	--	≥0.24	--	--	≥0.01	≥0.09	≥0.01	≥0.50	≥0.03	≥0.01	≥0.01	≥0.01

--: No data available to calculate the optimal threshold; --[a]: Breakdown in analysis due to small sample size

8.5.1. The specificity of optimal thresholds

Across all of the burglary samples examined in the present research, the placement of the decision threshold had a significant impact on the level of discrimination accuracy that could be achieved. As expected, the hit and false alarm rates always decreased as decision thresholds became stricter, that is, as higher levels of across-crime similarity were required in order to make a decision that crimes were linked. Such a finding clearly highlights the need to identify optimal decision thresholds for discrimination purposes so that an appropriate balance between hits and false alarms can be achieved within any given investigative setting.

To demonstrate how important it is to consider decision thresholds, inter-crime distances were focused on throughout the present research. As was the case with the AUCs for inter-crime distance, the optimal thresholds associated with these distances were found to be highly variable across the burglary samples, though they were always quite strict. Indeed, the thresholds varied from ≤1.90 km for residential burglars in Merseyside A to ≤5.20 km for commercial burglars in Merseyside A. The threshold patterns associated with inter-crime distances will now be discussed in relation to various aspects of the burglary data, as was done at the end of the previous section on behavioural consistency.

(a) Residential versus commercial burglary

As previously mentioned, a number of researchers have suggested that due to potential differences in target frequency and distribution, the spatial behaviour of serial burglars is effected by the type of burglary committed (e.g., Van Koppen & Jansen, 1998). The results from the present research have highlighted one possible difference already – residential burglaries tend to be characterised by slightly higher AUCs. As indicated by the maps in Figure 8.1, higher AUCs are found because the distributions of intra-offender and inter-offender distances in residential burglaries are thought to overlap to a small degree. In other words, compared to inter-crime distances in commercial burglary, intra-offender distances are assumed to be relatively short in residential burglary while inter-offender distances are assumed to be relatively long.

Based on the set of hypothetical maps presented in Figure 1.8, there is another difference that is expected as well, this time in relation to optimal decision thresholds. While the degree of distribution overlap between inter-crime distances determines what the AUCs will be for residential and commercial burglaries, the position of the distributions along the continuum of inter-crime distances is expected to determine the optimal decision threshold. Generally, as intra-offender or inter-offender distances get larger on average so to will the optimal decision threshold. This is exactly what would be expected to happen in the case of commercial burglary.

To examine whether such a pattern exists in the present research, the optimal decision thresholds for inter-crime distances are presented in Table 8.5, along with the average inter-crime distances for every burglary sample. From an examination of this information, inter-crime distances can be seen to directly relate to the leniency of the optimal decision threshold. Specifically, the results indicate that larger inter-crime distances are observed in commercial burglaries, and therefore more lenient decision thresholds need to be adopted for this type of crime. Indeed, from the 5 comparisons that can be made, all 5 of the decision thresholds are more lenient in the case of commercial burglary.

Table 8.5. Optimal decision thresholds in residential and commercial burglaries

	Residential		Commercial	
Sample	Distances[a]	Threshold	Distances	Threshold
London	15.70 1.80	≤4.60	-- --	--
Dorset	24.39 2.84	≤3.50	-- --	--
Oldham	3.16 1.16	≤2.00	4.13 1.37	≤2.50
Merseyside A	3.36 1.50	≤1.90	6.32 4.48	≤5.20
Merseyside B	10.73 1.54	≤2.60	7.45 3.31	≤3.00
Merseyside C	4.71 1.23	≤2.10	4.76 1.15	≤2.30
Merseyside D	7.81 1.53	≤2.20	8.70 1.31	≤2.30

[a] Average inter-crime distances are measured in kilometres, with the inter-crime distances for unlinked crime pairs presented above the inter-crime distances for linked crime pairs; --: No data available to calculate inter-crime distances or optimal decision thresholds

(b) Rural versus urban areas

In line with the argument presented above for why more lenient decision thresholds should be found in cases of commercial burglary, more lenient thresholds should also be found in cases where burglaries are committed in rural areas. Again, this is because of the assumption that rural burglaries are characterised by larger intra-offender and/or inter-offender distances due to the frequency and distribution of potential targets in these areas.

As before, because precise measures of population density in relation to each burglary location have not been collected in the present research, it is difficult to assess whether such differences really exist between rural and urban burglaries. However, to examine whether a relationship exists at a very general level the optimal decision thresholds for inter-crime distances have been presented in Table 8.6, along with the average inter-crime distances for every burglary sample. These have all been ordered with respect to the estimated population density of the samples from which they were derived.

From this table, it can be seen that very little evidence exists for such a relationship. Burglaries committed in areas with high population densities are not associated with either the shortest inter-crime distances or the strictest decision thresholds. Likewise, burglaries committed in areas with low population densities are not associated with the largest inter-crime distances or the most lenient decision thresholds. The reasons for not finding such a relationship are probably the same reasons for why a relationship was not found in relation to the AUCs. To reiterate, it could be that a relationship simply does not exist, it could be that the measure of population density used in the present research is not precise enough, or it could be that a relationship exists, just not with population density.

Table 8.6. Optimal decision thresholds in rural and urban areas

		Residential		Commercial	
Sample	Density[a]	Distances[b]	Threshold	Distances	Threshold
London	4611	15.70 1.80	≤4.60	-- --	--
Merseyside C	4053	4.71 1.23	≤2.10	4.76 1.15	≤2.30
Merseyside D	1875	7.81 1.53	≤2.20	8.70 1.31	≤2.30
Merseyside A	1875	3.36 1.50	≤1.90	6.32 4.48	≤5.20
Oldham	1551	3.16 1.16	≤2.00	4.13 1.37	≤2.50
Merseyside B	1467	10.73 1.54	≤2.60	7.45 3.31	≤3.00
Dorset	153	24.39 2.84	≤3.50	-- --	--

[a] The population densities for each police district are measured in persons/km^2; [b] Average inter-crime distances are measured in kilometres, with the inter-crime distances for unlinked crime pairs presented above the inter-crime distances for linked crime pairs; --: No data available to calculate the inter-crime distances or optimal decision threshold

(c) Force-wide versus divisional/district level burglaries

The level of precision with which data was collected in the present research may also be expected to relate to the optimal decision thresholds found for inter-crime distances. Specifically, burglaries collected at the force-wide level would be expected to have more lenient decision thresholds compared to burglaries collected at the divisional or district level. The fact that the area covered by an entire police force is greater than the area covered by a single police division or district suggests that inter-offender distances will be much larger in force-wide samples. In turn, these larger distances should result in larger optimal decision thresholds being required at the force-wide level.

Without exception, the inter-offender distances presented in Table 8.4 are found to be much larger for burglaries collected over larger areas. For example, inter-offender distances in the London and Dorset sample are 15.70 km and 24.39 km respectively, which is consistent with the fact that Dorset covers a much larger area than London does. Also in line with this reasoning, these distances are much larger than those found in Oldham and Merseyside. In these samples of burglary, which were collected at the divisional and district level, none of the inter-offender distances are larger than 10.73 km. As expected, these differences inter-crime distances are also reflected in the optimal decision thresholds. With the exception of inter-crime distances in Merseyside A, optimal thresholds are always found to be more lenient in the force-wide samples.

8.6. Validation of the ROC results

In reality, none of the results presented throughout this research would matter much if they could not be applied to crimes beyond those included in the burglary samples. To some extent, this issue was dealt with in the present research by validating all the logistic regression models using crimes that were not originally used during the model development phase. However, a more thorough test was undertaken in each chapter by running a series of external discrimination trials. To examine the validity of the ROC results across the burglary samples, the results from these trials will be focused on here.

Table 8.7 presents a summary of the results from the discrimination trials that were conducted in each of the previous chapters using inter-crime distances. Various pieces of information are included in this table. First, the values of *p*H and *p*FA obtained from the optimal operating points on each ROC graph are presented (the target values). Second, the average values of *p*H and *p*FA achieved across the discrimination trials in each chapter are presented (the observed values). Third, the standard deviations associated with the average values of *p*H and *p*FA are presented. Fourth, deviation scores are presented, which indicate the differences between the observed and target values.

This summary of the discrimination trials highlights a number of interesting patterns across the burglary samples examined in the present research. The first interesting point is that, at a very general level, the ROC graphs developed on each burglary sample seem to have a relatively high degree of predictive accuracy. In other words, across nearly all the samples, the target values of *p*H and *p*FA generally correspond to the values of *p*H and *p*FA observed in the discrimination trials. This is indicated by the fact that the deviation scores in Table 8.7 are typically very small, with only 4 of the 48 observed values having deviation scores that exceed +/−0.10.

A second interesting point is that a slightly higher degree of deviation is typically associated with *p*H. The average deviation scores for *p*H and *p*FA across all the discrimination trials are 0.06 and 0.04 respectively. This finding is probably due in part to the large discrepancies that exist between the number of unlinked and linked crime pairs. Because of the low frequency of linked crime pairs, fairly small variations in how accurately these crimes are linked will result in large variations in *p*H across the trials. In the case of unlinked crime pairs, even fairly large variations in how accurately these crimes are linked will result in small variations in *p*FA across the trials. For example, given 100 linked crime pairs and 1000 unlinked crime pairs, ten times more linking errors would have to be made when analysing the unlinked crime pairs to get the same proportion of linking errors.

A third interesting point is that the standard deviations associated with *p*H and *p*FA are typically smaller when the discrimination trials are run on larger samples. Specifically, the average standard deviation is 0.06 for the smaller samples and 0.02 for the larger samples. In line with the argument put forward in the preceding paragraph, this is likely to be a function of the relationship between the frequency of crime pairs available for analysis and the resulting variations in *p*H and *p*FA. As more unlinked and linked crime pairs become available for analysis, more linking errors will be needed to cause a substantial degree of variation in pH and pFA across the trials.

The implications of these findings are relatively clear. Essentially, the results from these trials suggest that the level of discrimination accuracy that is possible in yet to be observed serial burglaries is able to be determined to a large degree by ROC results established on solved serial burglaries from the same police jurisdiction. This seems to be especially true when dealing with unlinked crime pairs, though reasonable levels of predictive accuracy are also associated with linked crime pairs.

Table 8.7. Summary of the discrimination trials

	London	Dorset	Oldham residential	Oldham commercial	Merseyside residential				Merseyside commercial			
					A	B	C	D	A	B	C	D
Target pH	0.93	0.82	0.73	0.65	0.70	0.86	0.84	0.75	0.63	0.75	0.81	0.90
Target pFA	0.07	0.22	0.28	0.38	0.30	0.15	0.16	0.25	0.37	0.25	0.19	0.10
Avg. pH 1[a] (SD, deviation)	0.95 (0.03,+0.02)	0.79 (0.08,-0.03)	0.82 (0.14,+0.09)	0.72 (0.08,+0.07)	0.78 (0.10,+0.08)	0.81 (0.09,-0.05)	0.79 (0.01,-0.05)	0.80 (0.09,+0.05)	0.58 (0.30,-0.05)	0.78 (0.03,+0.03)	0.88 (0.06,+0.07)	0.82 (0.06,-0.08)
Avg. pH 2[b] (SD, deviation)	0.87 (0.02,-0.05)	0.70 (0.01,-0.12)	0.70 (0.01,-0.03)	0.87 (0.03,+0.18)	0.75 (0.03,+0.03)	0.78 (0.04,-0.08)	0.80 (0.03,-0.04)	0.78 (0.04,+0.03)	0.52 (0.02,-0.11)	0.77 (0.02,+0.02)	0.88 (0.02,+0.07)	0.83 (0.02,-0.07)
Avg. pFA 1[a] (SD, deviation)	0.07 (0.00,0.00)	0.19 (0.07,-0.03)	0.26 (0.01,-0.02)	0.31 (0.01,-0.07)	0.30 (0.05,0.00)	0.15 (0.02,0.00)	0.16 (0.01,0.00)	0.26 (0.03,+0.01)	0.45 (0.18,+0.08)	0.37 (0.01,+0.12)	0.28 (0.01,+0.09)	0.09 (0.02,-0.01)
Avg. pFA 2[b] (SD, deviation)	0.07 (0.00,0.00)	0.18 (0.00,-0.04)	0.19 (0.00,-0.09)	0.28 (0.00,-0.10)	0.35 (0.01,+0.05)	0.14 (0.00,-0.01)	0.17 (0.01,+0.01)	0.27 (0.01,+0.02)	0.42 (0.05,+0.05)	0.38 (0.01,+0.13)	0.29 (0.00,+0.10)	0.09 (0.01,-0.01)

[a] Refers to the average hit and false alarm rates obtained across discrimination trials using small sample sizes; [b] Refers to the average hit and false alarm rates obtained across discrimination trials using large sample sizes

8.7. Practical implications

The previous points raised in this discussion highlight the theoretical importance of the findings that emerged throughout this research. However, because the ROC results do appear to be fairly generalisable, the patterns of consistency and discrimination just discussed have a number of practical implications as well. The most obvious implication has to do with the possibility of using the ROC procedure as a diagnostic tool for carrying out comparative case analysis (CCA), where the goal is to determine whether the same offender has committed two or more crimes (Bennell & Canter, in press). Another practical implication has to do with using the ROC results as the basis for a legal argument of similar fact evidence (Alison, Bennell, Mokros & Ormerod, in press; Ormerod, 1999).

8.7.1. Using the ROC procedure as a tool for CCA

The patterns of consistency and discrimination that have been found in the present research are very directly related to CCA, especially given the ecologically valid nature of the data that has been used. Three implications in particular warrant further discussion. These relate to matters of priority, redundancy and specificity. Never before have such issues been considered in any depth.

(a) Prioritising behavioural domains

When carrying out CCA, the results presented here suggest that the use of different behavioural domains will result in drastically different levels of linking accuracy. Indeed, in certain police jurisdictions, this level of accuracy has been shown to vary from near perfect to around chance depending on the behavioural domain that is used. As a consequence, the results provide a way for the police to accurately and objectively prioritise the behaviours they use when conducting CCA in cases of serial burglary. This is something that has, perhaps surprisingly, never been proposed before.

The findings presented throughout this research suggest that inter-crime distances can provide a powerful and very simple way of linking burglaries. As a result, inter-crime distances may prove useful as a first filter when carrying out CCA. The goal of this first filter would be to reduce the number of potential links that initially need to be examined. Considering the current state of technology in the majority of modern day police forces, it would be feasible to draw on the results presented here and combine them with digital maps of police divisions to create likely 'linkage areas'. Crimes committed within certain distances of one another would be considered as potentially linked and given a high priority and then additional analytical techniques could be used to further reduce the number of false alarms.

It would be at this later stage that the other behavioural domains examined in the present research may prove of some use. However, even at this stage some ordering of domains could be carried out, by starting with target selection choices and entry behaviours for example. During these later stages, a number of other strategies would also have to be used in order to reduce the number of false alarms to a much more manageable number. Such strategies may include, but are not necessarily limited to, the use of signature analysis (Keppel, 2000)[8], the use of temporal analysis (Eskridge 1983), and the use of other police intelligence (Merry, 2000).

A number of people have recently made similar proposals (e.g., Grubin *et* al., 2001; House, 1997; Merry, 2000). Merry (2000), for example, has suggested that various strategies need to be combined in order to achieve maximum success in burglary investigations. The focus in Merry's work is on spatial and behavioural approaches to the investigation of burglaries, in a similar way to what has been done here, though he includes temporal analysis as another possible option. What the present research offers to proposals like this one is some empirical guidance in terms of how to go about ordering various investigative strategies from most to least effective so that maximum success can be achieved by the police as quickly as possible.

[8]Although later on in this chapter it is stated that behavioural signatures are expected to be relatively rare in high volume crimes such as burglary, certain offences examined throughout this research suggest that on some occasions they do exist. For example, one serial burglar repeatedly broke into occupied homes and told the occupants to stand in the bathtub while he searched and stole from the property.

(b) Reducing the redundancy of behavioural domains

Another implication to emerge from the present research in relation to CCA is that the use of many behavioural domains in combination may lead to insignificant increases in linking accuracy. The ROC procedure provides an accurate and objective procedure for discovering a way to achieve maximum predictive power in CCA using the fewest possible number of behavioural domains. This is incredibly important in the practical context given the fact that police officers can rarely spend a lot of time at a crime scene collecting and coding data.

The ROC procedure reduces the need to collect a great deal of information on a crime. Indeed, the results presented here suggest that collecting appropriate, possibly limited, information carefully may be more effective than collecting a great deal of information in the hope that some if it may turn out to be of value. Thus, instead of developing longer, more comprehensive linking pro forma's, which is currently the strategy adopted in many UK police forces, the method of analysis presented in this research opens up the possibility of findings ways to provide more manageable guidance that is just as effective. This can potentially be done simply by cutting out the unneeded redundancies in the behavioural features that are used in CCA.

(c) The need for specificity

A third implication to emerge from the present research in relation to CCA is that linking strategies will probably need to be area and crime type specific. This is the case for two reasons. The first reason is that the ordering of behavioural domains, in terms of their discriminatory power, varies depending on the sample being examined. For example, as the results in the present research reveal, just because inter-crime distances are associated with a high level of discrimination accuracy in Merseyside C, this does not mean that target selection choices are going to be associated with a similar level of accuracy in Merseyside A.

The second reason is that optimal decision thresholds are also sample specific. Thus, as the results in the present research also reveal, just because inter-crime distances are associated with the highest levels of discrimination accuracy for both Oldham residential and commercial burglary, this does not mean that the optimal decision thresholds associated with these inter-crime distances will be the same for both types of burglary in Oldham.

How specific linking systems should be remains an unanswered question. For example, the results presented in this research suggest that different linking strategies may certainly be needed at the level of different police districts. With the right data it would be possible to examine this issue further, by increasing the level of geographic precision to the beat and neighbourhood level. Given the present results, it would not be overly surprising to find that every police district requires a slightly different linking strategy for residential and commercial burglary. Indeed, it may only be at this level that a truly effective strategy for carrying out CCA can be attained. Whether this is practically feasible, however, is a matter that will have to be considered.

8.7.2. Using the ROC procedure as the basis for similar fact evidence

There have now been a number of court cases in Canada, the US and the UK where behavioural similarities have been drawn on to demonstrate that the same offender has committed two or more crimes (Ormerod, 1999). Typically, these cases involve serial rape or murder and the evidence takes the form of signature analysis. The goal of signature analysis is to link crimes together based on the expression of specific offence behaviours that are thought to be an enduring part of the offender who committed the crimes (Douglas & Munn, 1992). In Delaware v. Pennell (1991), for example, the trial court accepted testimony that 3 murders were the work of the same offender because each crime was characterised by similar ritualistic behaviours involving physical and sexual torture.

While there is some evidence for the existence of these behavioural signatures in violent interpersonal crimes such as rape and murder (e.g., Keppel, 2000), there are strong grounds for thinking they are likely to be rare and unlikely to be identifiable for very frequent crimes such as burglary (Canter, 2000). There is, therefore, some value in identifying the degree to which less specific behavioural features of an offence may help link that offence to others committed by the same offender. This would provide an alternative to signature analysis when attempting to establish similar fact evidence in cases that involve high volume crimes such as burglary.

The present research represents a first step towards finding such an alternative. Indeed, there is much to be said for simply creating databases of burglary behaviour that can be used to estimate in an objective and empirical fashion how unusual a burglary is in the UK. Beyond the creation of these databases, however, the present research provides an analytical framework that has the potential to identify the behaviours that should be used to develop similar fact evidence in cases serial burglary. It would appear from the results reported here, that instead of using crime scene behaviours to demonstrate that the same offender is responsible for a series of burglaries, the spatial aspect of the burglaries should be drawn on instead.

CHAPTER 9
FUTURE RESEARCH

9.1. Introduction

The analytical framework proposed in the present research opens up the possibility of conducting future research in a variety of different areas, each characterised by a different degree of specificity. One of these areas relates to research that could examine general psychological issues of broad importance across psychology. A second, slightly more specific area relates to research that could be carried out to examine a variety of tasks arising within the investigative context. Finally, research in the third and most specific area consists of further studies designed to unravel specific issues emerging from the present research. These three areas are not mutually exclusive, but possible avenues of future research in each area will be discussed in turn.

9.2. An examination of general psychological issues

The findings in the present research have many implications for understanding general psychological issues, beyond those arising specifically in the investigative context. First, the ROC procedure could be used to help understand behavioural consistency in the non-criminal context. Second, the ROC procedure could be used to help understand behavioural discrimination in the non-criminal context, in terms of how observers perceive peoples' behaviour.

9.2.1. Examining how people behave

Much of personality psychology over the last four decades has been concerned with identifying the conditions under which people behave in a consistent fashion (e.g., Bem & Allen, 1974; Chaplin, 1991; Emmons & Diener, 1986; Funder & Colvin, 1991; Zuckerman *et al.*, 1988). However, questions remain in this field over how to determine the extent to which a specific variable moderates consistency or how to measure the impact of one potential moderator compared to another (Tellegen, Kamp & Watson, 1982). The ROC procedure may have the potential to throw light on these issues. Specifically, it is proposed that the AUC could act as a measure of a variable's moderating power and also provide a way of comparing the relative impact of different moderator variables.

As a quick example of how this could be done, consider Funder and Colvin's (1991) study of moderator variables. Among other things, Funder and Colvin found that when behaviours are coded in terms of psychological themes rather than discrete behaviours, or when behaviours are rated as relatively stimulus free, levels of behavioural consistency tend to increase. As a result of these findings, Funder and Colvin concluded that 'level of abstraction' and 'level of specificity' moderate the degree to which consistency can be observed in peoples' behaviour. However, nothing in this study indicates how to calculate the degree of a moderator's influence, or how to compare the effect of two different moderators.

One possible way of answering these questions is to construct ROC curves for the various levels of each moderator variable. Essentially, if similarity measures were calculated across pairs of situations encountered by each participant in the different moderator groups, AUCs could be obtained and compared in the exact same way they have been here. It would be possible, for example, to compare an AUC calculated for behaviours coded at an abstract level to an AUC calculated for behaviours coded at a more discrete level. Furthermore, it would be possible to calculate AUCs for behaviours rated as stimulus free or stimulus specific, and to compare these with each other and with the AUCs calculated for the different levels of abstraction.

9.2.2. Examining how observers perceive behaviour

In addition to understanding the ways in which people express themselves across situations, personality psychologists also examine whether observers agree in their perceptions of this behaviour (e.g., Funder, 1982, 1995; Funder & West, 1993). Knowing exactly how to measure such agreement, however, is still somewhat of a problem. Interestingly, Ozer (1993) proposes a potential solution to this problem using the just noticeable difference measure from classical psychophysics, which was briefly mentioned in Chapter 2. For example, in response to a question such as, how different must two teachers judgements of a student be in order to qualify as a disagreement, Ozer states that:

> If the jnd of intelligence were a known quantity, then one might assert that a difference in ratings which failed to exceed the value of the jnd should

> not be counted as a consequential disagreement, since if two persons differed by this amount it would not be (by definition) a noticeable difference. (p. 741)

The risk that comes with such a solution, however, is that it may fall prey to the same problems faced by early psychophysicists. Essentially, the reliance on a just noticeable difference measure does not provide insight into the source of disagreement in the first place. For example, it may be the case that judges differ in their ability to discriminate between people, but it could equally be the case that different responses result from the use of different decision thresholds.

Again, the ROC procedure has the potential to throw light on these issues since it allows one to separate out measures of overall discrimination accuracy and the impact of adopting different decision thresholds. The procedure would allow one to determine whether two teachers, for example, fall on the same ROC curve when judging various differences between students, just at different operating points, or whether the teachers actually differ in their ability to identify meaningful differences. Amongst other things, these results could have substantial practical benefits in terms of providing appropriate training to teachers so that they can effectively evaluate students.

9.3. An examination of various investigative tasks

The present research also has implications for a variety of other tasks commonly faced by the police in the investigative context, beyond attempts to link serial crimes. Indeed, the ROC procedure could potentially be applied to any two-alternative investigative task. For example, consider tasks that require discriminations to be made between true or false allegations of rape, genuine or forged suicide notes, hostage negotiations that may result in success or failure, and so on. Each of these tasks requires at least two decisions to be made. The first relates to the sort of evidence that should be used to make the discrimination in the first place. The second relates to how much evidence should be available before the discrimination is actually made. Based on the results presented here, and a wealth of other research, the ROC procedure should at least be considered as a potentially useful way of resolving these issues.

As an example of how the ROC procedure could be applied to new investigative tasks, consider the problem faced by investigators when trying to predict where an offender lives based on the location of his crimes (Rossmo, 2000). The procedure for making these predictions using the ROC approach would be similar to the current task of linking crimes based on their inter-crime distances. In fact, there would really only be two important differences. First, instead of using inter-crime distances as the predictor variable, the relevant distances would be between homes and crimes. Second, instead of the criterion variable being whether two crimes are actually linked or unlinked, the relevant variable would be whether the correct offender's home is found or not. Essentially, the underlying assumption of the procedure is that most offenders will live relatively close to where they commit their crimes and, therefore, that there should be a strong relationship between the predictor and criterion variable.

A preliminary analysis of the Oldham residential burglary data, which includes both home and crime locations, suggests that the ROC procedure may provide a useful way of accomplishing this task. Indeed, a ROC curve constructed using the procedure employed throughout this research indicated that a home-to-crime distance of 3.00 km results in a high proportion of correct homes being found (pH=0.89) along with a relatively low proportion of incorrect homes (pFA=0.37). It must be remembered, however, that only the homes of known residential burglars in Oldham were included in this preliminary analysis. The results would be much worse if a different set of homes were used (e.g., Oldham offenders with any known previous conviction). One of the advantages of this approach is that unlike other methods for making these predictions, this method does require crimes to be linked initially.

9.4. A further examination of burglary behaviour

Despite the obvious implications that these results have for these other areas, more studies are also needed to unravel various issues that have emerged in the present research. Three areas of research in particular should receive further attention. The first area consists of studies designed to provide insight into why the results presented here emerge in the first place. The second area consists of further attempts to identify the precise conditions under which burglars exhibit behavioural consistency. The third area consists of attempts to understand and measure the impact that the present results could have in practical settings.

9.4.1. Gaining a better understanding of the present results

Due to the exploratory nature of the present research, some important questions remain unanswered. It is clear that serial burglars in the UK often exhibit offence behaviours in a consistent fashion across their crimes, and it is clear that these consistencies can lead to high levels of discrimination accuracy. However, why these consistencies emerge in the first place is still somewhat unknown. One of the weaknesses in the present research is the lack of detailed information available, about the offenders who committed the crimes, the crime themselves, and the geographic regions where the crimes have taken place.

There seems to be at least two major ways by which a better understanding of the present results could be achieved. The first way would be to conduct extensive face-to-face interviews with serial burglars. These could be done in the style of Cromwell, Olson and Avary (1991), Maguire (1982), or Wright and Decker (1994). The second way would be to gather as much detailed information as possible about the geographic regions where the offences take place. This could be done in the style of Baldwin and Bottoms (1976) or Hirschfield and Bowers (1997).

(a) Interviews with offenders

To gain a better understanding of the ROC results presented throughout this research, it seems likely that in-depth interviews with offenders will need to be carried out. Getting answers to detailed questions about what offenders do during the commission of their crimes, how they do it, why they do it, and perhaps whom they do it with, could help explain many of the results. For example, answers to such questions could provide insight into whether inconsistencies observed in burglary behaviour are primarily a result of internal factors such as maturation or learning, situational factors beyond the offender's control, group processes that take place as a result of co-offending, or simply the consequence of unreliable and inaccurate police data.

Equally clear is the fact that the ROC results could be used to guide some of this interview questioning. For example, the discovery that most offenders commit at least some of their crimes in a relatively well-defined offending territory would suggest this is something worth exploring in an interview setting. For example, it would be interesting to know whether offenders are aware of where other burglars are committing their crimes, whether a conscious decision is made to steer clear of these territories, and why offenders occasionally drift out of their established areas of criminal activity. Recent evidence does suggest that interviews with offenders can help illuminate some of the various psychological processes underlying these sorts of issues (e.g., Canter & Hodge, 2000; Canter & Shalev, 2000).

(b) Obtaining more detailed data

Obtaining more detailed information about the geographic areas where offences take place could also be another way to gain a deeper understanding of what the ROC results in the present research actually mean. Without such information it is difficult to know for certain why different results emerge across the samples. Why, for example, are inter-crime distances less consistent in Merseyside A compared to the other Merseyside districts? A range of information about the regions where offences take place may be useful for this purpose. This might include information about target density, demography, land use, crime rates, clear-up rates, social disorganisation factors, and so on. Collecting such information has already proved of some use in understanding certain aspects of burglary behaviour, such as the formation of hot spots (Hirschfield & Bowers, 1997) and repeat victimisation (Johnson, Bowers & Hirschfield, 1997).

Indeed, having now provided an analytical framework for examining behavioural consistency and discrimination in serial burglary, and having now tested that framework across different crime types and geographic areas, a more focused and in-depth research project would be extremely useful. For example, a productive line of future research would be to collect a representative sample of serial burglaries committed in a smaller area than the areas focused on here, and to collect as much detailed information as possible about the offenders, the offences, and the area. It would be useful to choose an area and a time period that is relevant to a specific investigative team within a particular police force, and to carry out a comprehensive analysis of serial burglary behaviour that could then be of use to those investigators.

9.4.2. Identifying the conditions under which burglars are consistent

In addition to conducting research that would provide us with a better understanding of the present results, it will also be important to carry out further research to identify the conditions under which burglars exhibit the highest levels of behavioural consistency. Not only would this sort of research further our understanding of criminal activity generally, and the factors that effect its expression, this research would also improve our ability to discriminate between crimes committed by different offenders.

A wide variety of moderators are probably worthy of study, but three in particular will be focused on here. The first potential moderator has to do with how behavioural consistency is defined. The second has to do with how behavioural consistency is measured. The third has to do with what offence behaviours are observed.

(a) Alternative definitions of behavioural consistency

There are many ways in which behavioural consistency can be defined, beyond the definitions used in the present research. Currently, consistency is defined at a relatively precise level, in terms of whether discrete actions are exhibited in a stable fashion across crimes. However, this is not the only level at which consistency can be found. Indeed, as mentioned earlier in this chapter, there is already some evidence in the non-criminal literature to suggest that behavioural consistency becomes more apparent when consistency is defined at a much higher level of abstraction.

In their studies of behavioural consistency, Funder and Colvin (1991) found a relatively high degree of consistency at the level of psychological themes but not at the level of discrete behaviours, that is, when behaviours were coded as an expression of aggression, rather than as punches, yells and tantrums. Altering the definition of consistency in this way may result in higher levels of observed consistency in the criminal context as well. Thus, one could examine whether burglars are consistent in terms of their level of skill, for example, regardless of what skilled behaviours the burglars actually exhibit. This approach may even be more advantageous in the criminal context than it is in the non-criminal context, because coding behaviours at a more abstract level may minimise some of the problems encountered when attempting to measure the consistency of burglary behaviour (e.g., the fact that a burglar's behaviour depends to a large extent on what is possible in the immediate offending environment).

(b) Alternative methods for measuring behavioural consistency

There are also a number of different ways in which behavioural consistency can be measured, and there is no way of knowing whether the methods of measurement adopted in the present research result in the highest levels of consistency. It would be useful to alter at least two aspects of the measurement procedure in future research to determine what effect this will have on observed levels of consistency. One aspect is related to the types of similarity scores used and the other relates to the method of statistical analysis.

The similarity score used to measure spatial consistency in the present research consists of Euclidean distances computed between each and every crime. However, there are a variety of other measures that could be used for this purpose that may result in quite different findings. Consider, for example, two burglaries that are close together when measured by their Euclidean distance but far apart when measured by a shortest route distance. Using the Euclidean distance, the two crimes would be perceived as the work of the same offender, though this might not be the case if shortest route distances were used instead. Likewise, the similarity scores used to measure the consistency of behavioural domains such as target selection choices consist of Jaccard's similarity coefficient. A number of other measures are available for use and could prove more effective, though the choice may be somewhat limited by the dichotomous nature of the data (Liebetrau, 1983).

It may also be worthwhile testing other statistical methods, beyond the logistic regression models used in the present research. Indeed, a number of studies have demonstrated that discrimination accuracy is at least somewhat dependent on the statistical method used. For example, Steadman *et al.* (2000) found that decision tree analysis resulted in higher discrimination accuracy compared to logistic regression analysis when predicting whether offenders would exhibit violent behaviour upon release from an institution. One of the reasons for this improvement was that, in contrast to the regression approach, decision tree analysis does not assume a single solution fits equally well to all offenders. Given that AUCs and optimal thresholds are found to be sample specific, decision tree analysis may prove to be a more suitable approach for examining consistency and discrimination in serial burglary.

(c) Alternative sets of behaviours

While the behavioural domains examined in the present research exhaust the variety of behaviours exhibited by burglars, the specific behaviours contained within these domains will have an enormous impact on the observed levels of consistency and discrimination. Therefore, it would be worthwhile in future research to examine very closely whether certain burglary behaviours result in higher levels of behavioural consistency and discrimination.

A starting point for this research might be to consider the frequencies of offence behaviours. There are a variety of reasons why behavioural frequencies should be related to observed levels of consistency in the criminal context. Canter (2000) draws attention to the fact that virtually every offender must exhibit certain offence behaviours because these behaviours are what define the crime in question. If one were to study behavioural consistency in serial rape, for example, it would be unproductive to include vaginal penetration in a 'sex domain' since every offender in the sample must exhibit this behaviour.

Alternatively, in a similar way to what Funder and Colvin (1991) found, it may be productive to focus on burglary behaviours that appear less dependent on specific situational stimuli. Just as the various behavioural domains examined in the present research seem to be ordered based on how much they rely on situational stimuli, the specific behaviours within each of these domains may also be ordered in a similar way.

It should also be pointed out that examining these sorts of issues in the future would open up the whole question of weighting behaviours. Essentially, the idea is to assign weights to behaviours in accordance with how important they are for discriminate between crimes. On the one hand, taking into account the frequency or specificity of behaviour may prove to be a productive way of empirically deriving these weights, and assigning such weights to behaviours may improve ones ability to link serial burglaries. On the other hand, it may turn out that weighting behaviours adds little in the way of increased predictive power. This is certainly the case sometimes, as Dawes (1979) showed in his tests of weighted and unweighted linear regression models.

9.4.3. Putting the results into practice

It would also be potentially useful to put the results presented here into practice, to see if they improve the performance of investigators in the field. Research in this area could cover at least two areas. The first area would involve testing investigators to see how reliably and accurately they make discrimination decisions, and then examining the extent to which the ROC procedure influences their performance. The second area would examine various ways in which optimal decision thresholds could be set to reflect, more accurately, the goals within specific investigative settings.

(a) Increasing discrimination performance in practical settings

One obvious direction to take the present research is to determine the extent to which the ROC procedure can be used to improve decisions made by investigators. There is a certain risk associated with the use of actuarial decision aids in the investigative context, however, in that it is often assumed they will automatically enhance performance (Snook, Canter & Bennell, 2002). Therefore, the first stage in this exploration must be to examine how reliable and accurate unaided investigators are in their attempts to discriminate between crimes committed by different offenders. Only then should an attempt be made to measure the extent to which decision support effects this performance.

The ROC approach is ideally suited for such an examination. Indeed, the majority of studies using ROC analysis have been carried out with this purpose in mind (e.g., Getty, Pickett, D'Orsi & Swets, 1988; Seltzer *et al.*, 1997). At its most basic level, the approach would be relatively simple. Investigators could be presented with a series of crime pairs along with the corresponding evidence and they would be required to state how likely it is that each crime pair is linked. These results could then be pooled to form an average ROC curve indicating the level of unaided accuracy (Swets & Pickett, 1982). Investigators could then be provided with decision support, provided in a variety of forms, and the process could be repeated resulting in a comparable ROC curve.

While there is no evidence available at present to suggest that discrimination accuracy would increase under such conditions, it does seem likely that improvements could be made. Even if investigators can identify the most appropriate behavioural domains to draw on for discrimination purposes it is unlikely they could identify optimal decision thresholds, and choosing an appropriate threshold has been shown to dramatically effect discrimination accuracy. If investigators were able to select optimal decision thresholds, at the very least the ROC procedure could ensure that investigators make linking decisions in a reliable fashion over time. This alone could be very beneficial.

(b) Improving the utility of discrimination decisions

A second line of research would be to examine possible ways of identifying more appropriate decision thresholds for specific investigative situations. Instead of assuming that an optimal threshold is one that allows the maximum number of hits to be made along with the minimum number of false alarms, it might be possible to assign specific costs and benefits to the various decision outcomes. For example, the cost of missing a pair of crimes committed by the same offender may be viewed as more serious in cases of residential versus commercial burglary due to the personal nature of residential crimes.

The evaluation of costs and benefits in the investigative context may depend on a whole host of factors, including the seriousness of the crime being investigated, the resources a police force has available, guidelines set out in government policy, the quality of police data, and so on. Taking all these factors into account for the purpose of assigning costs and benefits to various decision outcomes will undoubtedly be difficult, especially because human rights and lives are at stake. However, by carefully carrying out such cost-benefit analyses, there is the potential to fine-tune the results presented throughout this research to meet the particular demands of specific investigative situations, in a similar way to what is being done in a variety of other diagnostic settings (e.g., Schwartz, Dans & Kinosian, 1988).

9.5. Getting the police on board

None of this research could have taken place without the assistance provided by numerous police forces and police personnel. However, to ensure that better research is done in the future a number of changes must take place. Each of these changes requires the police to modify how they deal with burglary offences, and therefore it is important that the police see some value in research of this type. Some of these changes have already been implemented since the start of this research project, though not necessarily because of it, while others are being proposed here as a result of experiences gained through carrying out the research. On the basis of these experiences there are at least three issues that need to be addressed – data quality, data storage and data access.

(a) Issues concerning data quality

While some of the police forces participating in the present research are amongst the leading forces in the UK when it comes to collecting quality data, most forces still use collection procedures that will most likely result in unreliable and inaccurate information. To improve these procedures a variety of things can be done. To start with, behavioural coding sheets could be constructed so that the same information is considered at each and every crime scene, with space provided so that the investigating officer can record any other significant observations. It would be crucial for the behaviours included on these sheets to be clearly defined and as objective as possible so that different coders can accurately interpret what each behaviour means.

In order to collect the highest quality data, training must also be provided to all those involved in the collection process. Specifically, investigating officers would have to be trained on how to use the behavioural coding sheets, including a thorough discussion of what each behaviour means. Ideally, a relatively small group of officers would be responsible for all data collection in order to increase the reliability with which data is coded. If many different individuals are involved in the collection process, as will probably be the case, periodic checks should be made to ensure that data is being collected in a reliable fashion. If problem areas do arise, these should be brought to the attention of the data collectors as quickly as possible so that they can be resolved.

Many of the police officers involved in the present research also voiced concerns over how long it takes them to collect behavioural data from a crime scene. Most indicated that it was not possible, given the range of other responsibilities they had, for them to do a thorough and accurate job when carrying out this task. In part, this problem might be resolved by the creation of new positions in police forces where one individual could be responsible for data collection in a particular area. Indeed, some police forces have adopted such a strategy and claim that it does result in higher quality data.

An alternative, yet equally promising approach, would be to make the collection procedure as automated as possible. For example, some North American police forces have resorted to installing their behavioural coding sheets onto laptop computers in order to decrease the amount of time required for inputting crime scene data. Not only would this strategy reduce the time it takes to collect behavioural data from a crime scene, it would also make it easy to download that data onto a mainframe computer for storage and later analysis.

(b) Issues concerning data storage

Even if data could be easily downloaded onto a computer, changes to current data storage procedures would be required in order to carry out high quality research. Indeed, the way data is currently stored in many UK police forces is not conducive to analysing behavioural information. Again, a variety of things can be done to improve this situation, especially if some of the points discussed above are adopted. While the free-text portion of the collection procedure is useful for a variety of reasons, and should certainly be stored, most forms of analysis undertaken by crime analysts and external researchers require numerical data. Because of this, it would probably be more effective if all data were stored in dichotomous format as well, so as to indicate the presence or absence of specific crime scene behaviours.

In addition to making changing to the format in which data is stored, ones ability to analyse data would be also much improved if this data were stored on a single computer system. Currently, multiple computer systems usually need to be accessed in order to obtain the information required, especially if that information relates to different aspects of an offence. For example, many police forces in the UK currently store offence information, offender information, victim information, and forensic information on different computer systems.

Related to this issue, databases containing information about these various aspects should, at the very least, be able to relate easily to one another and to various analytical software packages such as digital mapping programs. At present, this is not the case in many UK police forces, which makes it very difficult to determine whether any relationships exist between these different pieces of information. For example, researchers working within the Merseyside Police Service have suggested that it would in fact be quite difficult to match up data that is collected about a particular burglary offence with the data collected about the offender responsible for that offence. This is the case because these two different pieces of information are stored in two separate databases that require data to be stored in slightly different ways.

(c) Issues concerning data access

Even if issues of data quality and storage are successfully sorted out, getting access to police data will ultimately dictate what sort of research can be carried out, as will getting access to the various software packages required to productively analyse that data. In the present research, access to police data was gained in part by ensuring that the project was viewed as a collaborative effort. Indeed, throughout the present research, much time was spent discussing with the participating police forces what the research could eventually offer them. In return for these discussions, and regular research updates, various personnel in each of the participating police forces were always willing to provide extra assistance and advice.

In the future, even more effort must be put towards involving police personnel in every stage of the research process. It is important to point out that developing these relationships is not only about getting access to data. In fact, the degree of police involvement in research of this type will influence if, when, and how the research is put into practice. Indeed, numerous studies have shown that decision aides are only ever adopted and relied upon in applied settings if the eventual decision-makers are actively involved in the aid's development (Adelman, 1982; Kaplan, Reneau & Whitecotton, 2001). Getting a police force to actively use the proposed approach for a time, so that they can carry out an evaluation study, is exactly what is needed in order to determine whether the present research can have a significant impact on how serial burglaries are actually investigated.

REFERENCES

Adelman, L. (1982). Involving users in the development of decision-analytic aids: The principal factor in successful implementation. Journal of the Operational Research Society, 33, 333-342.

Alison, L.J., Snook, B. and Stein, K.L. (2001). Unobtrusive measurement: Using police information for forensic research. Qualitative Research, 1, 241-254.

Alison, L.J., Bennell, C., Mokros, A. and Ormerod, D. (in press). The personality paradox in offender profiling: A theoretical review of the processes involved in deriving background characteristics form crime scene actions. Psychology, Public Policy and Law.

Baldwin, J. and Bottoms, A.E. (1976). The Urban Criminal. London: Tavistock Press.

Barker, M. (2000). The criminal range of small-town burglars. In D.V. Canter and L.J. Alison (Eds.), Profiling Property Crimes (pp. 57-74). Aldershot: Dartmouth.

Bem, D.J. and Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behaviour. Psychological Review, 81, 506-520.

Bennell, C. and Canter, D.V. (in press). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. Science & Justice.

Bennell, C., Alison, L.J., Stein, K., Alison, E. and Canter, D.V. (2001). Sexual offences against children as the abusive exploitation of conventional adult-child relationships. Journal of Social and Personal Relationships, 18, 155-171.

Bennett, T. and Wright, R.T. (1984). Burglars on Burglary: Prevention and the Offender. Aldershot: Gowan.

Blackwell, H.R. (1952). Studies of psychophysical methods for measuring visual thresholds. Journal of the Optical Society of America, 42, 606-616.

Brantingham, P.L. and Brantingham, P.J. (1991). Environmental Criminology. Prospect Heights, Il: Wavel and Press.

Brown, B. and Harris, P. (1989). Residential burglary victimisation: Reactions to the invasion of primary territory. Journal of Environmental Psychology, 9, 119-132.

Butler, G. (1994). Commercial burglary: What offenders say. In M. Gill (Ed.), Crime at Work: Studies in Security and Crime Prevention (pp. 29-41). Leicester: Perpetuity Press.

Canter, D.V. (1995). The psychology of offender profiling. In R. Bull and D. Carson (Eds.), Handbook of Psychology in Legal Contexts (pp. 343-355). Chichester: John Wiley and Sons.

Canter, D.V. (2000). Offender profiling and criminal differentiation. Legal and Criminological Psychology, 5, 23-46.

Canter, D.V., and Fritzon, K. (1998). Differentiating arsonists: A model of firesetting actions and characteristics. Legal and Criminological Psychology, 3, 73-96.

Canter, D.V. and Heritage, R. (1990). A multivariate model of sexual offence behaviour: Developments in offender profiling I. Journal of Forensic Psychiatry, 1, 185-217.

Canter, D.V. and Heritage, R. (1991). A Facet Approach to Offender Profiling. Final Report to the Home Office. University of Surrey, Department of Psychology, Offender Profiling Research Unit.

Canter, D.V. and Hodge, S. (2000). Criminals' mental maps. In L.S. Turnbull, E.H. Hendrix and B.D. Dent (eds.), Atlas of Crime: Mapping the Criminal Landscape (pp. 184-191). Phoenix, AZ: Oryx Press.

Canter, D.V. and Shalev, K. (December, 2000). Putting crime in its place: Psychological processes in crime site selection. Paper presented at the National Institute of Justice Crime Mapping Research Centre Conference, San Diego, CA.

Canter, D.V., Hughes, D. and Kirby, S. (1998). Paedophilia: Pathology, criminality, or both? The development of a multivariate model of offence behaviour in child sexual abuse. Journal of Forensic Psychiatry, 9, 532-555.

Canter, S.B. Sun, C.C., Tortolero-Luna, G., Richards-Kortum, R. and Follen, M. (1999). A comparison of C/B ratios from studies using receiver operating curve analysis. Journal of Clinical Epidemiology, 9, 885-892.

Capone, D. and Nichols, W.J. (1976). Urban structure and criminal mobility. American Behavioural Scientist, 20, 199-201.

Carter, G. and Polger, P. (1986). A 20-year Summary of National Weather Service Verification Results for Temperature and Precipitation. Technical Memorandum NWS FCST 31. Washington D.C.: National Oceanic and Atmospheric Administration.

Chaplin, W.F. (1991). The next generation of moderator research in personality psychology. Journal of Personality, 59, 143-178.

Chatterjee, S. and Price, B. (1997). Regression Analysis by Example. New York: John Wiley and Sons.

Clark, R.V. and Hough, J.M. (Eds.) (1980). The Effectiveness of Policing. Farnborough, Hants: Glower.

Coleman, C. and Moynihan, J. (1996). Understanding Crime Data: Haunted by the Dark Figure. Buckingham: Open University Press.

Craik, M. and Patrick, A. (1994). Linking serial offences. Policing, 10, 181-187.

Cromwell, P.F., Olsen, J.N. and Avery, D.W. (1991). Breaking and Entering: An Ethnographic Analysis of Burglary. Newbury, CA: Sage.

Davies, A. (1992). Rapist's behaviour: A three-aspect model as a basis for analysis and the identification of serial crime. Forensic Science International, 55, 173-194.

Davies, A. (1996). Locating the Stranger Rapist. London: Police Research Group.

Davies, N. (1999, March 18). Watching the detectives: How the police cheat in fight against crime. The Guardian, p 12.

Dawes, R.M. (1979). The robust beauty of improper linear models in decision-making. American Psychologist, 34, 571-582.

Delaware v. Pennell, 584 A. 2d. 513 (Del. Super. Ct. 1989).

Douglas, J.E. and Munn, C. (1992). Violent crime scene analysis: Modus operandi, signature and staging. FBI Law Enforcement Bulletin, 61, 1-10.

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

Eliason, S.R. (1993). Maximum Likelihood Estimation: Logic and Practice. Thousand Oaks, CA: Sage.

Emmons, R.A. and Diener, E. (1986). Situation selection as a moderator of response consistency and stability. Journal of Personality and Social Psychology, 51, 1013-1019.

Eskridge, C.W. (1983). Prediction of burglary: A research note. Journal of Criminal Justice, 11, 67-75.

Farrington, D.P. and Lambert, S. (1992). The Feasibility of a Statistical Approach to Offender Profiling: Burglary and Violence in Nottinghamshire. London: Home Office.

Fechner, G.T. (1966). Elements of Psychophysics (H.E. Adler, Trans.). New York: Holt, Rinehart and Winston. (original work published 1860).

Funder, D.C. (1982). On the accuracy of dispositional versus situational attributions. Social Cognition, 1, 205-222.

Funder, D.C. (1995). On the accuracy of personality judgement: A realistic approach. Psychological Review, 102, 652-670.

Funder, D.C. and Colvin, C.R. (1991). Explorations in behavioural consistency: Properties of persons, situations and behaviours. Journal of Personality and Social Psychology, 5, 773-794.

Funder, D.C. and West, S.G. (1993). Consensus, self-other agreement, and accuracy in personality judgement: An introduction. Journal of Personality, 64, 457-476.

Getty, D.M. Pickett, R.M., D'Orsi, C.J. and Swets, J.A. (1988). Enhanced interpretation of diagnostic images. Investigative Radiology, 23, 240-252.

Getty, D.J., Seltzer, S.E., Tempany, C.M.C., Pickett, R.M., Swets, J.A. and McNeil, B.J. (1997). Prostrate cancer: Relative effects of demographic, clinical, histologic, and MR imaging variables on the accuracy of staging. Genitourinary Radiology, 204, 471-479.

Gigerenzer, G. and Murray, D.J. (1987). Cognition as Intuitive Statistics. Hillsdale, NJ. Erlbaum.

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

Green, E.J., Booth, C.E., and Biderman, M.D. (1976). Cluster analysis of burglary MO's. Journal of Police Science and Administration, 4, 382-387.

Green, D.M. and Swets, J.A. (1988). Signal Detection Theory and Psychophysics. Los Altos, CA: Peninsula Publishing.

Greiner, M., Pfeiffer, D., and Smith, R.D. (2002). Principles and practical application of the receiver-operating characteristics analysis for diagnostic tests. Preventative Veterinary Medicine, 45, 23-41.

Grubin, D., Kelly, P. and Brunsdon, C. (2001). Linking Serious Sexual Assaults through Behaviour. London: Police Research Group.

Guttman, L. (1941). An outline of the statistical theory of prediction. In P. Horst (Ed.), The Prediction of Personal Adjustment (pp. 264-268). New York: Social Science Research Council.

Guttman, L. (1981). Efficacy coefficients for differences among averages. In I. Borg (Ed.), Multidimensional Data Representations: When and Why (pp. 20-46). Ann Arbor, MI: Mathesis Press.

Guttman, L. (1988). Eta, Disco, Odisco and F. Psychometrika, 53, 393-405.

Hanley, J.A. and McNeil, B.J. (1982). The meaning and use of the area under the receiver operating characteristic curve. Radiology, 143, 29-36.

Hanley, J.A. and McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology, 148, 839-843.

Hawkins, D.M. (Ed.). (1982). Topics in Applied Multivariate Analysis. Cambridge, MA: Cambridge University Press.

Heal, K. and Morris, P. (1981). Crime Control and Policing. Home Office Research Study No. 67. London: HMSO.

Hertz, J., Krogh, A., and Palmer, R.G. (1991). Introduction to the Theory of Neural Computation. Reading, MA: Addison-Wesley.

Hilden, J. (1991). The area under the ROC curve and its competitors. Medical Decision Making, 11, 95-101.

Hirschfield, A. and Bowers, K.J. (1997). The development of a social, demographic and land use profiler for areas of high crime. British Journal of Criminology, 37, 103-120.

Hosmer, D.W. and Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley and Sons.

House, J.C. (1997). Towards a practical application of offender profiling: The RNC's criminal suspect prioritisation system. In J.L. Jackson and D.A. Bekerian (Eds.), Offender Profiling: Theory, Research and Practice (pp. 177-190). New York: John Wiley and Sons.

Iversen, G.R. and Norpoth, H. (1987). Analysis of Variance. Beverly Hills, CA: Sage Publications.

Johnson, S.D., Bowers, K. and Hirschfield, A. (1997). New insights into the spatial and temporal distribution of repeat victimisation. British Journal of Criminology, 37, 224-241.

Kaplan, S.E., Reneau, J.H. and Whitecotton, S. (2001). The effects of predictive ability information, locus of control, and decision-maker involvement on decision aid reliance. Journal of Behavioural Decision Making, 14, 35-50.

Keppel, R.D. (2000). Signature murders: A report of the 1984 Cranbrook, British Columbia cases. Journal of the Forensic Sciences, 45, 508-511.

Keppel, R.D. and Weis, J.G. (1993). Improving the Investigation of Violent Crime: The Homicide Investigation and Tracking System. Washington D.C.: U.S. Department of Justice.

Lachenbruch, P. (1975). Discriminant Analysis. New York: Hafner.

Lee, R.M. (2000). Unobtrusive Methods in Social Research. Milton Keynes: Open University Press.

Lewis-Beck, M.S. (1980). Applied Regression: An Introduction. London: Sage Publications.

Liebetrau, A.M. (1983). Measures of Association. Beverly Hills, CA: Sage Publications.

Lo, J.Y., Baker, J.A., Kornguth, P.J., Iglehart, J.D., and Floyd, C.E. (1997). Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features. Radiology, 203, 159-163.

Maguire, M. (1980). The impact of burglary on victims. The British Journal of Criminology, 20, 261-275.

Maguire, M. (1982). Burglary in a Dwelling. London: Heinemann.

Maltz, M.D., Gordon, A.C. and Friedman, W. (1990). Understanding and using information about crime. In Maltz, M.D., Gordon, A.C. and Friedman, W.

(Eds.), Mapping Crime in its Community: Event Geography Analysis (pp. 44-58). New York: Springer-Verlag.

McClelland, D.C. (1984). Is personality consistent? In D.C. McClelland (Ed.), Motives, Personality and Society (pp. 185-211). New York: Praeger.

McGinty, J. and Williams, T. (2001). Regional Trends (Number 36). London: The Stationary Office.

McNicol, D. (1972). A Primer of Signal Detection Theory. London: Allen and Unwin.

Merry, S. (2000). Crime analysis: Principles for analysing everyday serial crime. In D.V. Canter and L.J. Alison (Eds.), Profiling Property Crimes (pp. 297-318). Aldershot: Dartmouth.

Metz, C.E. (1978). Basic principles of ROC analysis. Seminars in Nuclear Medicine, 8, 283-298.

Metz, C.E., Hermann, B.A. and Shen, J.H. (1988). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously distributed data. Medical Decision Making, 17, 1033-1053.

Metz, C.E., Starr, S.J. and Lusted, L.B. (1997). Quantitative evaluation of visual detection performances in medicine: ROC analysis and determination of diagnostic benefit. In G.A. Hay (Ed.), Medical Images (pp. 220-241). London: Wiley.

Mischel, W. and Peake, P.K. (1982). Beyond deja vu in the search for cross-situational consistency. Psychological Review, 89, 730-755.

Neyman, J. and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. Philosophic Transactions of the Royal Society of London, 231, 289-337.

Obuchowski, N.A. (1994). Computing sample size for receiver operating characteristic studies. Investigative Radiology, 29, 238-243.

Ormerod, D. (1999). Criminal profiling: Trial by judge and jury, not criminal psychologist. In D.V. Canter and L.J. Alison (Eds.), Profiling in Policy and Practice (pp. 209-261). Aldershot: Dartmouth.

Ozer, D.J. (1993). Classical psychophysics and the assessment of agreement and accuracy in judgments of personality. Journal of Personality, 61, 739-767.

Partin, A.W. Kattan, M.W. Subong, E.N.P., Walsh, P.C., Wojno, K.J. Desterling, J.E., Scardino, P.T. and Pearson, J.D. (1997). Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological cancer. Journal of the American Medical Association, 277, 1445-1451.

Peterson, W.W., Birdsall, T.G. and Fox, W.C. (1954). The theory of signal detectability. IRE Professional Group on Information Theory, PGIT-4, 171-212.

Rengert, G.F. and Wasilchick, J. (2000). Suburban Burglary: A Tale of Two Suburbs. Springfield, IL: Charles Thomas Publishers.

Rose, D. (1995). Psychophysical methods. In G.M. Breakwell, S. Hammond and C. Fife-Schaw (Eds.), Research Methods in Psychology (pp. 140-159). London: Sage Publications.

Rossmo, D.K. (2000). Geographic Profiling. Boca Ration, FL: Chemical Rubber Company.

Salfati, C.G. and Canter, D.V. (1999). Differentiating stranger murders: Profiling offender characteristics from behavioural styles. Behavioral Sciences and the Law, 17, 391-406.

Scheffe, H. (1959). The Analysis of Variance. New York: Wiley.

Schwartz, J.S., Dans, P.E. and Kinosian, B.P. (1988). Human immunodeficiency virus test evaluation, performance, and use: Proposals to make good tests better. Journal of the American Medical Association, 259, 2574-2579.

Seltzer, S.E. Getty, D.J., Tempany, C.M.C., Pickett, R.M., Schnall, M.D., McNeil, B.J. and Swets, J.A. (1997). Staging prostate cancer with MR imaging: A combined radiologist-computer system. Radiology, 202, 219-226.

Shoda, Y. (1999). Behavioural expressions of a personality system: Generation and perception of behavioural signatures. In D. Cervone and Y. Shoda (Eds.), The Coherence of Personality: Social-cognitive Bases of Consistency, Variability, and Organisation (pp. 155-184). New York: The Guilford Press.

Shoda, Y., Mischel, W. and Wright, J. (1994). Intra-individual stability in the organisation and patterning of behaviour: Incorporating psychological situations into the idiographic analysis of personality. Journal of Personality and Social Psychology, 67, 674-687.

Skinner, B.F. (1966). The Behaviour of Organisms: An Experimental Analysis. New York: Appleton-Century-Crofts.

Snook, B., Canter, D.V. and Bennell, C. (2002). Predicting the home location of serial offenders: Comparing the accuracy of human judges to a geographic profiling system. Behavioral Sciences and the Law, 20, 109-188.

Steadman, H.J., Silver, E., Monahan, J., Appelbaum, P., Robbins, P., Mulvey, E., Grisso, T., Roth, L. and Banks, S. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. Law and Human Behavior, 24, 83-100.

Swanson, C.R., Chamelin, N.C. and Territo, L. (1996). Criminal Investigation. New York: McGraw-Hill.

Swets, J.A. (Ed.) (1964). Signal Detection and Recognition by Human Observers: Collected Readings. New York: John Wiley and Sons.

Swets, J.A. (1973). The relative operating characteristic in psychology. Science, 182, 990-1000.

Swets, J.A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. Psychological Bulletin, 99, 100-117.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. Science, 240, 1285-1293.

Swets, J.A. (1992). The science of choosing the right decision threshold in high-stake diagnostics. American Psychologist, 47, 522-532.

Swets, J.A. (1996). Signal Detection Theory and ROC Analysis in Psychology and Diagnostics. Mahwah, NJ: Erlbaum.

Swets, J.A. and Pickett, R.M. (1982). Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York: Academic Press.

Swets, J.A., Dawes, R.M. and Monahan, J. (2000a). Psychological science can improve diagnostic decisions. Psychological Science in the Public Interest, 1, 1-26.

Swets, J.A., Dawes, R.M. and Monahan, J. (2000b). Better decisions through science. Scientific American, October, 82-87.

Tabachnick, B.G. and Fidell, L.S. (1996). Using Multivariate Statistics. New York: Harper Collins Publishers.

Tellegen, A., Kamp, J. and Watson, D. (1982). Recognising individual differences in predictive structure. Psychological Review, 89, 95-105.

Thurstone, L.L. (1927). A law of comparative judgement. Psychological Review, 34, 273-286.

Turvey, B. (2000). Modus operandi. In G. Knupfer, P. Saukko, and J. Seigal (Eds.), Encyclopedia of Forensic Sciences (pp. 1116-1121). London: Academic Press

Van Koppen, P.J. and Jansen, R.W.J. (1998). The road to the robbery: Travel patterns in commercial robberies. British Journal of Criminology, 38, 230-246.

Walsh, D. (1980). Break-ins: Burglary from Private Houses. London: Constable.

Walsh, D. (1986). Heavy Business: Commercial Burglary and Robbery. London: Routledge & Kegan Paul.

Welkowitz, J., Ewen, R.B. and Cohen, J. (1982). Introductory Statistics for the Behavioral Sciences (3rd Edition). New York: Academic Press.

Wiersma, E. (1996). Commercial burglary in the Netherlands: Reasoning decision makers? International Journal of Risk, Security and Crime Prevention, 1, 217-225.

Wiggins, J.S. (1973). Personality and Prediction: Principles of Personality Assessment. Reading, MA: Addison-Wesley.

Wiles, P. and Costello, A. (2000). The Road to Nowhere: The Evidence for Travelling Criminals. London: Home Office Policing and Reducing Crime Unit.

Wilson, M., Canter, D.V., Jack, K. and Butterworth, D. (1997). The Psychology of Rape Investigations: A Study in Police Decision-making. ESRC Report. Unpublished Manuscript.

Wright, R.T. and Decker, S.H. (1994). Burglars on the Job: Streetlife and Residential Break-ins. Boston: Northeastern University Press.

Zuckerman, M., Koestner, R., DeBoy, T., Garcia, K.T., Maresca, B.C. and Sartoris, J.M. (1988). To predict some of the people some of the time: A re-examination of the moderator variable approach in personality theory. Journal of Personality and Social Psychology, 54, 1006-1019.

APPENDIX A
CALCULATING PROBABILITIES

In order to understand the various analytical methods used and discussed throughout the present research, it is necessary to understand how a variety of probabilities are calculated. The purpose of this appendix is to provide a brief overview of how to calculate these probabilities and to make clear where each fits into the present research. A more thorough discussion of these issues can be found in Swets *et al.* (2000a), where most of the information presented here was collected.

A.1. The data upon which the probabilities are based

Most of the probability calculations carried out in the present research are based on the same source of data. This data is usually presented in the form of a two-by-two contingency table of predictions versus reality, also called a classification matrix, a confusion matrix, or a decision matrix (see Table A1). The columns of this two-by-table typically relate to reality – whether a pair of crimes has been committed by the same offender (actually linked) or different offenders (actually unlinked). The rows of this two-by-two table typically relate to predictions – whether a pair of crimes is predicted to have been committed by the same offender (predicted linked) or by different offenders (predicted unlinked). After making predictions, the frequencies of cases falling into each of the four cells are labelled *a*, *b*, *c* and *d*, and refer to the frequency of hits, false alarms, misses and correct rejections respectively (Swets *et al.*, 2000a).

Table A1. A two-by-two contingency table of reality versus predictions

		Reality: Linked	Reality: Unlinked	
Prediction:	Linked	*a* hit	*b* false alarm	*a*+*b*
	Unlinked	*c* miss	*d* correct rejection	*c*+*d*
		a+*c*	*b*+*d*	*a*+*b*+*c*+*d*=*N*

A.2. Conditional probabilities

The most important probability calculation to understand in the present research is referred to as a conditional probability calculation. The probability is termed conditional because it refers to the probability of making a particular decision conditional upon the existence of a particular reality (Swets *et al.*, 2000a). Thus, the probability of predicting that a pair of crimes is linked given that they are in fact linked is a conditional probability, the probability of a hit (*p*H) in the present context. Of course, there are three other conditional probabilities that can occur. These are referred to as the probability of a false alarm (*p*FA), the probability of a miss (*p*M), and the probability of a correct rejection (*p*CR). The reason these probabilities are the most important ones in the present research is because they form the basis of the ROC curve.

Conditional probabilities can be easily estimated using the frequencies (*a*, *b*, *c* and *d*) in the above table. Dividing any of these frequencies by a row or column total results in a different conditional probability (Swets *et al.*, 2000a). Thus, the conditional probability of a hit, *p*(actually linked | predicted linked), is calculated by dividing *a* by *a*+*c*. The conditional probability of a false alarm, *p*(actually unlinked | predicted linked), is calculated by dividing *b* by *b*+*d*. The conditional probability of a miss, *p*(actually linked | predicted unlinked), is calculated by dividing *c* by *a*+*c*. Lastly, the conditional probability of a correct rejection, *p*(actually unlinked | predicted unlinked), is calculated by dividing *d* by *b*+*d*.

A.3. Prior probabilities

A second probability calculation that is important to understand in the present research is referred to as a prior probability calculation. Prior probabilities refer to the probability of certain realities occurring. In the present context, there are two prior probabilities, the probability of an unlinked crime pair actually occurring in a particular police jurisdiction and the probability of a linked crime pair actually occurring. Prior probabilities in the present context are important because, ideally, they should effect where the decision threshold is placed. For example, if the probability of a linked crime pair actually occurring is relatively low, say 0.01, the decision threshold should be placed much lower than if the probability of a linked crime pair actually occurring is relatively high, say 0.50. Since the occurrence of linked crime pairs in the present research is always very low, it should not be surprising that optimal decision thresholds will also be low.

Prior probabilities can also be easily estimated using the frequencies in Table A1. Dividing column sums by the total sample size gives the prior probabilities of the two alternative realities (Swets *et al.*, 2000a). Thus, the prior probability of an unlinked crime pair is calculated by dividing *b*+*d* by *N*. The prior probability of a linked pair is calculated by dividing *a*+*c* by *N*.

A.4. Inverse probabilities

The last probability calculation that is important to understand in the present research is referred to as an inverse (or Bayesian) probability calculation. In a sense, the inverse probability is the complete opposite of the conditional probability. Unlike a conditional probability, which is the probability of making a particular decision conditional upon a particular reality, the inverse probability is the probability of a particular reality existing given a particular piece or amount of evidence. Inverse probabilities are important in the present research because these are the probabilities that all predictions are based on. As Swets and his colleagues (2000a) point out, these are the probabilities that the "...[logistic regression analysis] supplies for diagnosis and forms the continuum of evidence along which a decision threshold is set to permit a binary, positive or negative, decision" (p. 26).

Given a particular piece, or amount of evidence, denoted e, inverse probabilities can be calculated using Bayes theorem (Swets *et al.*, 2000a). The evidence in the present case consists of various across crime similarity scores that can range in value. When crimes have in fact been committed by the same offender, the formula for calculating p(actually linked) given a particular amount of evidence e is:

$$\frac{p(e \mid \text{actually linked}) \times p(\text{actually linked})}{p(e)}$$

where $p(e)$ for actually linked crimes equals:

$$[p(e \mid \text{actually linked}) \times p(\text{actually linked})] + [p(e \mid \text{actually linked}) \times p(\text{actually linked})]$$

When crimes have in fact been committed by different offenders, the formula for calculating p(actually unlinked) given a particular amount of evidence e is:

$$\frac{p(e \mid \text{actually unlinked}) \times p(\text{actually unlinked})}{p(e)}$$

APPENDIX B
ROC CALCULATIONS

In order to understand the ROC analyses presented throughout this research, it may be helpful to present a variety of calculations required for carrying out the procedure. The main purpose of this appendix is to provide a brief overview of these calculations and to make clear where each fits into the present research. A second purpose of this appendix is to present ROC calculations not used in the present research, due to limitations of the data, which may be used in the investigative context to explore similar issues. A more thorough discussion of these issues can be found in Swets (1996) or Greiner *et al.* (2000), where most of the information presented here was collected.

B.1. Non-parametric versus parametric ROCs

Empirical ROCs can take a non-parametric or parametric form. If one does not wish to assume that the underlying signal-noise distributions are normally distributed with equal variance, than a non-parametric ROC is the best option (Swets, 1996). To construct a non-parametric ROC, one simply plots *p*H against *p*FA across various decision thresholds and connects the resulting ROC points. This is the procedure used throughout the present research, though it was done automatically using SPSS.

If one does wish to assume that the underlying signal-noise distributions are normally distributed with equal variance, than a parametric ROC will be the best option (Swets, 1996). To construct a parametric ROC, one needs to calculate a number of additional parameters, denoted A and B in most of the literature (Metz, 1978). To calculate the parameter A, let x_0 and x_1 denote the mean values of some similarity score for unlinked and linked crime pairs respectively, and s_0 and s_1 denote the standard deviations of that similarity score. So long as $x_0 < x_1$ than $A=(x_1-x_0)/s_1$, which is the standardised mean difference between the two groups. To calculate the parameter B, take the ratio of the two standard deviations, such that $B=s_0/s_1$. A parametric ROC is then constructed by supplying a ROC program (e.g., ROCKIT: Metz, Hermann & Shen, 1988) with these two ROC parameters.

B.2. Calculating the area under a ROC curve

There are a variety of ways to calculate the AUC. The choice of which procedure to use is largely based on whether one is dealing with a non-parametric or parametric ROC. In the case of non-parametric ROCs, there are at least two procedures that can be used. These are commonly referred to as the trapezoidal rule and the Wilcoxin area estimate (used in the present research). In the case of parametric ROCs, there is only one procedure typically used. This procedure takes into account the A and B parameters defined above, and is referred to by the symbol A_z.

B.2.1. The trapezoidal rule

The simplest procedure to use for calculating the AUC of a non-parametric ROC is the trapezoidal rule. One simply takes a ROC curve consisting of numerous points connected by straight lines, draws a vertical line from each ROC point to the x-axis, and sums the resulting trapezoidal and triangular areas. The area of each trapezium is equal to half the distance between the parallel sides multiplied by the sum of the two parallel sides and the area of a triangle is equal to half the length of its base multiplied by its altitude. The general problem with this method is that it consistently underestimates AUCs based on a smoothed ROC curve because connecting ROC points with straight lines reduces some of the true area.

B.2.2. The Wilcoxin area estimate

Another non-parametric method for calculating the AUC is called the Wilcoxin area estimate (Hanley & McNeil, 1982). The formula for calculating the Wilcoxin area estimate is:

$$\frac{n_0 n_1 - \mathrm{U}}{n_0 n_1}$$

where n_0 and n_1 denote the number of unlinked and linked crime pairs respectively, and the term $U=R-\frac{1}{2}n_0(n_0+1)$, where R is the rank sum of the unlinked crime pairs.

To test whether the Wilcoxin area estimate is significantly greater than chance, we can form a null hypothesis where the expected value of the rank sum $E(R)=\frac{1}{2}n_0(n+1)$, resulting in $U=\frac{1}{2}(n_0n_1)$ and AUC=0.50. The null hypothesis can be tested using the test statistic z, which is equal to:

$$\frac{(\mathrm{R}-\mathrm{E(R)})}{\sqrt{\mathrm{var(R)}}}$$

The variance of R can be estimated by $\mathrm{var(R)}=(n_0n_1s^2)/n$, where s^2 is the sample variance of the combined ranks for both groups.

It is also possible to calculate a standard error, SE(W), for the Wilcoxin area estimate (Hanley & McNeil, 1982). The formula for doing so is:

$$\mathrm{SE}(W)=\sqrt{\frac{\mathrm{AUC}(1-\mathrm{AUC})+(n_1-1)(Q_1-\mathrm{AUC}^2)+(n_0-1)(Q_2-\mathrm{AUC}^2)}{n_0n_1}}$$

where

$$Q_1=\frac{\mathrm{AUC}}{(2-\mathrm{AUC})}$$

and

$$Q_2=\frac{2(\mathrm{AUC}^2)}{(1+\mathrm{AUC})}$$

B.2.3. The parametric AUC

The parametric approach for calculating AUC considers the parameters A and B defined above. In addition, it considers $\Phi(z)$, which is the cumulative frequency distribution function of the standard normal distribution (Obuchowski, 1994). The formula for calculating this measure is:

$$\Phi\left(\frac{A}{\sqrt{1+B^2}}\right)$$

To test whether this area estimate is significantly greater than chance, we can form a null hypothesis where we expect $x_0=x_1$. In this case, A would equal 0 and the AUC would equal 0.50. The general problem with this method is that the distributional assumptions required for its use are rarely ever satisfied. Consequently, the Wilcoxin area estimate is often the favoured method (Greiner *et al.*, 2000).

B.3. Comparison of two ROC curves

If d is the difference between the AUCs associated with two ROC curves, values of d close to zero indicate the two curves have the same level of accuracy. We can establish the standard error of d, SE(d), using the following formula:

$$\mathrm{SE}(d) = \sqrt{\mathrm{var}(\mathrm{AUC}_1) + \mathrm{var}(\mathrm{AUC}_2) - 2r\mathrm{SE}(\mathrm{AUC}_1)\mathrm{SE}(\mathrm{AUC}_2)}$$

where $\mathrm{var}(\mathrm{AUC}_i)=Q_{1i}+Q_{2i}-2\mathrm{AUC}_i^2$ (Q_1 and Q_2 defined as above) is an estimate of the variance of the AUC for test i (i=1,2), r is an estimate of the correlation between the values of the two data sets, and $\mathrm{SE}(\mathrm{AUC}_i)$ equals:

$$\sqrt{\mathrm{var}(\mathrm{AUC}_i)}$$

(Hanley & McNeil, 1983). We can then construct confidence limits around *d.* If the confidence interval for the differences between the two AUCs includes 0, we can conclude that there is no significant difference between the two ROC curves.

APPENDIX C
INSTRUCTIONS FOR USING *S-LINK*

S-LINK is a computer program developed to calculate spatial similarity scores in the present research. To use *S-LINK*, one must have access to geo-coded x and y coordinates, which indicate the geographic location of any given site (e.g., crime site, home location, work place, etc.). These x and y coordinates can be of any length, though the length will effect the accuracy of all *S-LINK* output. Typically, the x and y coordinates are either 6 digits or 7 digits long. The longer the coordinates, the more accurate *S-LINK* will be.

To run *S-LINK*, all x and y coordinates must be in a Microsoft Excel comma delimited file (e.g., 'London.csv'). The x and y coordinates should take up the first two columns of the worksheet and there should be no other information included in the file beyond these coordinates. The .csv file containing these coordinates must then be transferred to the folder containing *S-LINK*. At this point, the .csv file must be renamed 'input.txt' so that it can be read by *S-LINK*. Renaming of files is most often done in the MS-DOS environment using the rename command (e.g., 'rename London.csv input.txt'). The file 'in.txt' is then read into *S-LINK* by changing the directory in MS-DOS to *S-LINK* and typing 'javaBrent'. This command begins the processing of calculations and sends all output to a file named 'output.txt'. This file can be located in the folder containing *S-LINK* and opened in Microsoft Excel as a .csv file for viewing.

The only limitation with this analysis at present has to do with the number of coordinates that can be initially entered into the program. The problem occurs not because of the number of coordinates entered per se, but because of the number of crime pairs that are output as a result. At present, the output from *S-LINK* cannot exceed 60,000 crime pairs, though this should be sufficient for most purposes. It is important to point out that this problem has nothing to do with *S-LINK* itself. Instead, the limiting factor is the number of rows that are available to hold data in Microsoft Excel. This currently stands at 60,000 rows.

APPENDIX D
INSTRUCTIONS FOR USING *B-LINK*

B-LINK is a computer program developed to calculate behavioural similarity scores in the present research. To use *B-LINK*, one must have access to information about offence behaviours that can be converted into dichotomous data (i.e., 0/1 values). Typically, a value of 0 would indicate a particular behaviour was absent and a value of 1 would indicate a particular behaviour was present. The output from *B-LINK* consists of a variety of similarity measures calculated between each and every crime, including measures based on Jaccard's coefficient, Yule's Q and Pearson's phi. However, the measures based on Jaccard's coefficient are the only ones used in the present research.

To run *B-LINK*, all dichotomous data must be in a Microsoft Excel comma delimited file (e.g., 'entry.csv'). The first column of the worksheet should specify in numeric form what crime the data corresponds to (e.g., 1-1 would refer to offender 1 – crime 1). The next *n* columns must consist of the dichotomous data from behaviour 1 to behaviour *n*. The first row of the worksheet must consist of a series of labels. In the first column, the label should read 'offence'. In the next *n* columns, the labels should read 'b1' to 'b_n'. The .csv file containing this data must then be transferred to the folder containing *B-LINK*. At this point, the .csv file must be renamed 'in.txt' so that it can be read by *B-LINK*. Renaming of files is most often done in the MS-DOS environment using the rename command (e.g., 'rename entry.csv in.txt'). The file 'in.txt' is then read into *B-LINK* by changing the directory in MS-DOS to *B-LINK* and typing 'java CraigApp >out'. This command begins the processing of calculations and sends all output to a file named 'out.txt'. This file can be located in the folder containing *B-LINK* and can be opened in Microsoft Excel as a .csv file for viewing.

As is the case with S-*LINK*, the only limitation with this analysis at present has to do with the number of crimes that can be initially entered into the program. Again, the problem occurs not because of the number of crimes entered per se, but because of the number of crime pairs that are output as a result. At present, the output from *B-LINK* cannot exceed 60,000 crime pairs. This problem has nothing to do with *B-LINK* itself. Instead, the limiting factor is the number of rows that are available to hold data in Microsoft Excel, which currently stands at 60,000 rows.

APPENDIX E

VARIABLE LIST FOR DORSET RESIDENTIAL BURGLARY

This variable list was developed as a means of coding residential burglary behaviours from the Dorset Crime Database. This list contains both the abbreviated variable name and more descriptive variable labels, both of which correspond to the Dorset data files found on the attached CD-ROM.

General information:
Offid: Offender ID number
Crimenum: Crime number
Year: Year of crime

Spatial behaviour:
Offencex: X coordinate
Offencey: Y coordinate

Target selection choices:
Wellmain: Well maintained
Avgmain: Average maintained
Poormain: Poorly maintained
Ndetach: Non-detached dwelling
Detach: Detached dwelling
Midterr: Mid-terraced dwelling
Endterr: End-terraced dwelling
Enclose: Access enclosed
Unocc: Unoccupied
Alarm: Alarm
Seclight: Security light

Internal behaviours:
Unsrch: Untidy search
Intsrch: Intrusive search
Nosrch: No search
Multsrch: Multiple rooms searched
Privsrch: Private rooms searched
Maldam: Malicious damage
Offsec: Secured dwelling
Defec: Defecated/urinated
Lookout: Used lookout
Foodcon: Consumed food

Entry behaviours:
Force: Access using force
Insec: Access using insecurity
Glass: Access by breaking glass
Ground: Access from ground
Upper: Access from upper level
Front: Access from front
Rear: Access from rear
Side: Access from side
Window: Access using window
Door: Access using door
Tosc: Brought instrument to scene
Frsc: Used instrument from scene

Property stolen:
Wallet: Wallet stolen
Fuel: Petrol stolen
Cig: Tobacco stolen
Tool: Tools stolen
Buildsup: Building supplies stolen
Comp: Computer hardware stolen
Compgame: Computer games stolen
Gun: Ammunition/firearms stolen
Vehicle: Vehicle stolen
Sport: Sports equipment stolen
Clothes: Clothing stolen
Cycle: Bicycle stolen
Video: TV/video stolen
Cd: CDs/cassettes stolen
Case: Case stolen
Photo: Photo equipment stolen
Drug: Drugs/pharmaceuticals stolen
Offeq: Office equipment stolen
Doc: Documents stolen
Domitem: Domestic items stolen
Key: Keys stolen
Food: Food/alcohol stolen
Clock: Clock/watch stolen
Av: Audio visual equipment stolen
Jewel: Jewellery stolen
Cash: Cash stolen
Fiveplus: Five items or more stolen
Nothing: No items stolen

APPENDIX F

VARIABLE LIST FOR OLDHAM RESIDENTIAL BURGLARY

This variable list was developed as a means of coding residential burglary behaviours from the Manchester Crime Pattern Analysis system. This list contains both the abbreviated variable name and more descriptive variable labels, both of which correspond to the Oldham residential burglary data files found on the attached CD-ROM.

General information:
Offnum: Actual offender number
Offid: Offender ID number
Crimno: Crime number

Spatial behaviour:
Offencex: X coordinate
Offencey: Y coordinate

Target selection choices:
Detach: Detached dwelling
Nondetac: Non-detached dwelling
Terraced: Terraced dwelling
House: House
Flat: Flat
Alarm: Alarm
Unocc: Unoccupied

Property stolen:
Domitem: Domestic items stolen
Medicine: Medicine/drugs stolen
Firearm: Firearms stolen
Bldsupp: Building supplies stolen
Furn: Furniture stolen
Shop: Shop/pub fittings stolen
Av: Audio visual equipment stolen
Clothes: Clothing stolen
Comp: Computer equipment stolen
Food: Food/drinks stolen
Doc: Documents stolen
Jewel: Jewellery stolen
Offeq: Office equipment stolen
Key: Keys/locks/safes stolen
Wallet: Wallet/handbag stolen
Tool: Tools stolen
Book: Books/magazines stolen
Sporteq: Sporting equipment stolen
Car: Car stolen
Cash: Cash stolen
Nil: No items stolen

Entry behaviours:
Entdoor: Access using door
Entwind: Access using window
Entgrd: Access from ground
Entside: Access from side
Entfront: Access from front
Entrear: Access through rear
Byalarm: Bypassed alarm
Frclock: Forced lock
Smash: Smashed glass
Force: Bodily force
Insecure: Access using insecurity
Remove: Removed glass
Frsc: Used instrument from scene
Tosc: Brought instrument to scene

APPENDIX G

VARIABLE LIST FOR OLDHAM COMMERCIAL BURGLARY

This variable list was developed as a means of coding residential burglary behaviours from the Manchester Crime Pattern Analysis system. This list contains both the abbreviated variable name and more descriptive variable labels, both of which correspond to the Oldham commercial burglary data files found on the attached CD-ROM.

General information:
Offid: Offender ID number
Crimeno: Crime number
Year: Year of crime

Spatial behaviour:
Offencex: X coordinate
Offencey: Y coordinate

Target selection choices:
Garage: Garage
Pub: Pub
Newagent: Newsagent
Rest: Restaurant
Market: Market
Petrol: Petrol station
Factory: Factory
Mill: Mill
Othshop: Other shop
Daycare: Day care
Takeaway: Takeaway
Hairdr: Hairdressers
Commcent: Community centre
Hotel: Hotel
Shed: Shed
School: School
Depstore: Department store
Jewstore: Jewellery store
Sportcent: Sports centre
Socclub: Social club
Surgery: Surgery
Church: Church
Carpark: Car park
Other: Other
Alarm: Alarm
Unocc: Unoccupied

Property stolen:
Domitems: Domestic items stolen
Bldsupp: Building supplies stolen
Furn: Furniture stolen
Shop: Shop/pub fittings stolen
Av: Audio visual equipment stolen
Clothes: Clothing stolen
Compeq: Computer hardware stolen
Mach: Machines stolen
Food: Food/drink stolen
Doc: Documents stolen
Jewel: Jewellery stolen
Offeq: Office equipment stolen
Key: Keys/locks/safes stolen
Wallet: Wallet/handbag stolen
Bike: Bike stolen
Tool: Tools stolen
Photoeq: Photo equipment stolen
Sporteq: Sporting equipment stolen
Tobacco: Tobacco items stolen
Cash: Cash stolen
Nil: No items stolen

Entry behaviours:
Entdoor: Access using door
Entwind: Access using window
Entgrd: Access from ground
Entup: Access from upper level
Entrear: Access from rear
Climb: Access by climbing
Byalarm: Bypassed alarm
Force: Forced lock
Remove: Removed glass
Frsc: Used instrument from scene

APPENDIX H

VARIABLE LIST FOR MERSEYSIDE RESIDENTIAL BURGLARY

This variable list was developed as a means of coding residential burglary behaviours from the Merseyside Integrated Criminal Justice System. This list contains both the abbreviated variable name and more descriptive variable labels, both of which correspond to the Merseyside residential burglary data files found on the attached CD-ROM.

General information:
Offid: Offender ID number
Beat: Crime beat
Beatnum: Beat number
Year: Year of crime

Spatial behaviour:
Offencex: X coordinate
Offencey: Y coordinate

Target selection choices:
House: House
Flat: Flat
Shed: Shed
Garage: Garage
Oldage: Old age home
Other: Other
Unocc: Unoccupied

Property stolen:
Av: Audio visual equipment stolen
Cd: CD stolen
Jew: Jewellery stolen
Intox: Intoxicating substances stolen
Clothtoi: Clothing/toiletries stolen
Cashgen: Cash generator stolen
Carr: Carrier stolen
Personal: Personal items stolen
Cash: Cash stolen
Wallet: Wallet stolen
Smelec: Small electrical items stolen
Lgelec: Large electrical items stolen
Tv: TV stolen
Vcr: VCR stolen
Camera: Photo equipment stolen
Comp: Computer equipment stolen
Vidcam: Video camera stolen
Stereo: Stereo equipment stolen
Sport: Sports equipment stolen

Internal behaviours:
Multsrch: Multiple rooms searched
Privsrch: Private rooms searched
Unsrch: Untidy search
Extsrch: Extended search
Maldam: Malicious damage
Drop: Dropped stolen items
Usefac: Used facilities
Forcar: Forensic carelessness
Secprem: Secured premises
Frcintdr: Forced interior door

Entry behaviours:
Entf: Access from front
Entr: Access from rear
Ents: Access from side
Entdoor: Access using door
Entwind: Access using window
Entceil: Access using ceiling
Byalarm: Bypassed alarm
Climb: Access by climbing
Inst: Access using instrument
Insec: Access through insecurity
Smglass: Access by smashing glass
Remglass: Access by removing glass
Bodforc: Bodily force
Key: Access using key
Con: Access using confidence trick
Entgr: Access from ground
Entup: Access from upper level

APPENDIX I

VARIABLE LIST FOR MERSEYSIDE COMMERCIAL BURGLARY

This variable list was developed as a means of coding residential burglary behaviours from the Merseyside Integrated Criminal Justice System. This list contains both the abbreviated variable name and more descriptive variable labels, both of which correspond to the Merseyside commercial burglary data files found on the attached CD-ROM.

General information:
Offid: Offender ID number
Beat: Crime beat
Beatnum: Beat number
Year: Year of crime

Spatial behaviour:
Offencex: X coordinate
Offencey: Y coordinate

Target selection choices:
Drugstor: Pharmacy
Grocstore: Grocery store
Newagent: Newsagent
Othersto: Other store
Pub: Pub
Rest: Restaurant
Takeaway: Takeaway
Arcade: Arcade
School: School
Colluni: College/university
Church: Church
Servstat: Service station
Hosp: Hospital
Hotel: Hotel
Sports: Sports club
Othclub: Other club
Factory: Factory
Office: Office

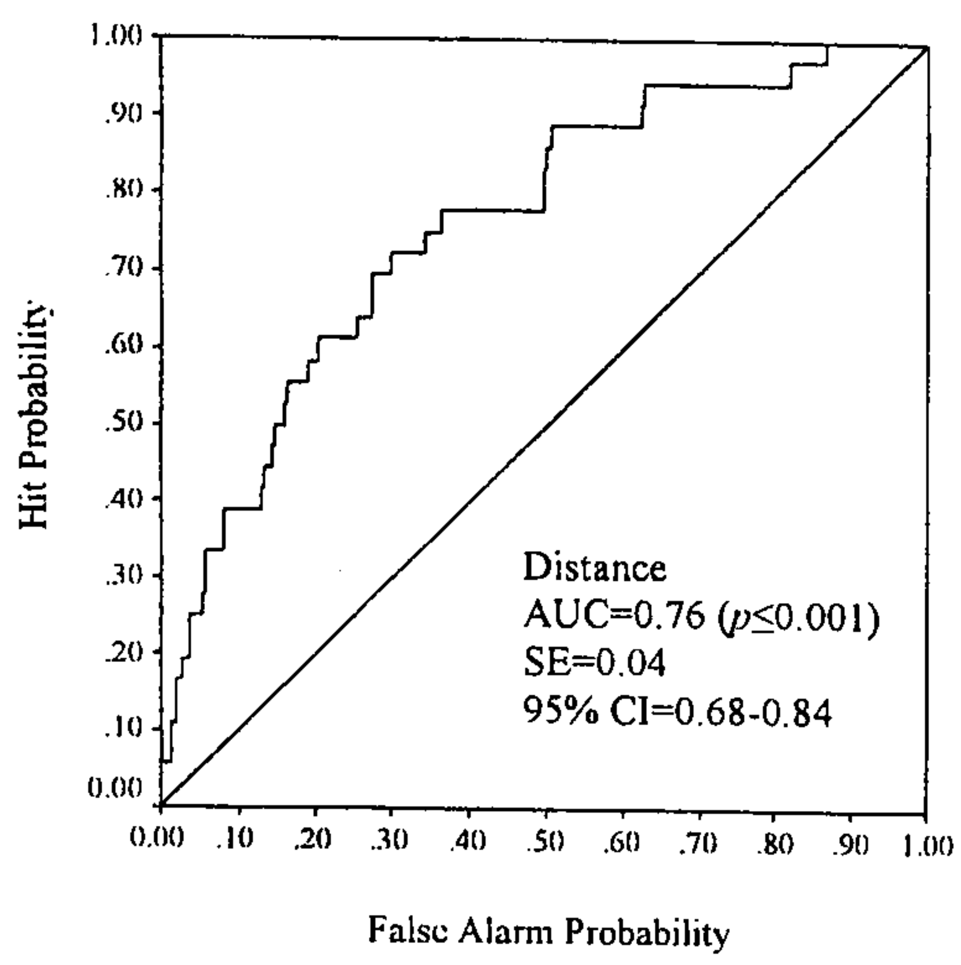
Property stolen:
Av: Audio visual equipment stolen
Cd: CDs stolen
Jew: Jewellery stolen
Intox: Intoxicating substances stolen
Clothes: Clothes stolen
Furn: Furniture stolen
Cashinst: Cash instrument stolen
Lugg: Luggage stolen
Comp: Computer equipment stolen
Cash: Cash stolen
Wallet: Wallet stolen
Smelec: Small electric items stolen
Lgelec: Large electric items stolen
Food: Food stolen
Kitch: Kitchen equipment stolen
Photo: Photo equipment stolen
Teleq: Telephone equipment stolen
Toilet: Toiletries stolen

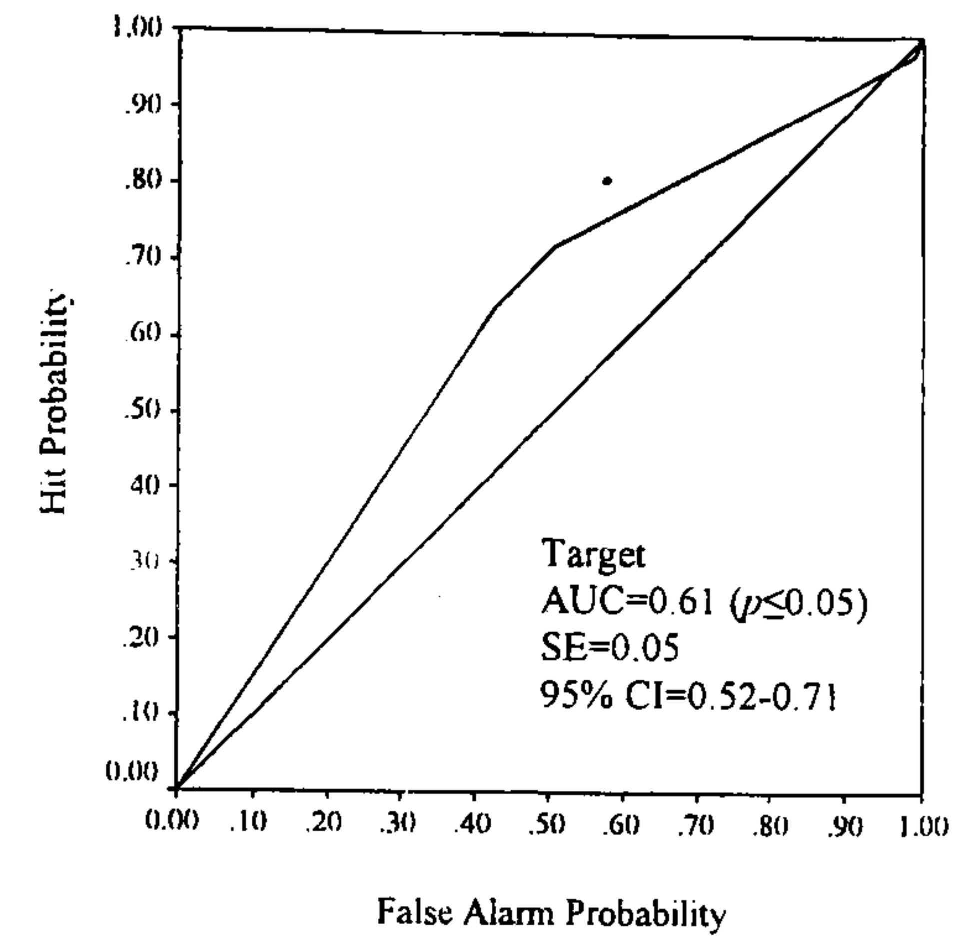
Internal behaviours:
Forcar: Forensic carelessness
Maldam: Malicious damage
Drop: Dropped stolen items
Carr: Used carrier
Intdam: Internal damage
Unsrch: Untidy search
Nosrch: No search
Extsrch: Extended search
Multsrch: Multiple rooms search
Exasent: Exit as entered

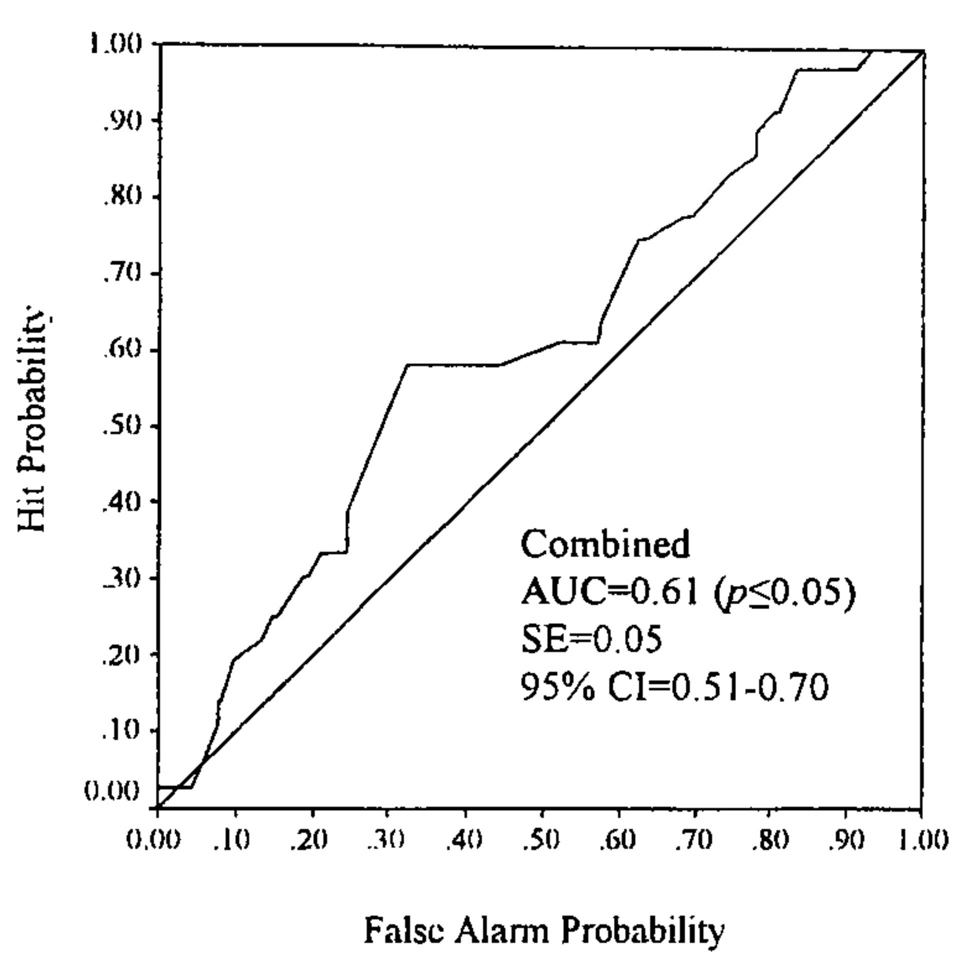
Entry behaviours:
Smashed: Smashed glass
Removed: Removing glass
Ewall: Access through wall
Eroof: Access through roof
Edoor: Access through door
Ewind: Access through window
Eside: Access from side
Efront: Access from front
Erear: Access from rear
Ebase: Access from basement
Adjbld: Access through next door
Byalarm: Bypassed alarm
Climb: Access by climbing
Insec: Access through insecurity
Sneak: Sneaked
Frcshutt: Forced shutters open
Inst: Access using instrument

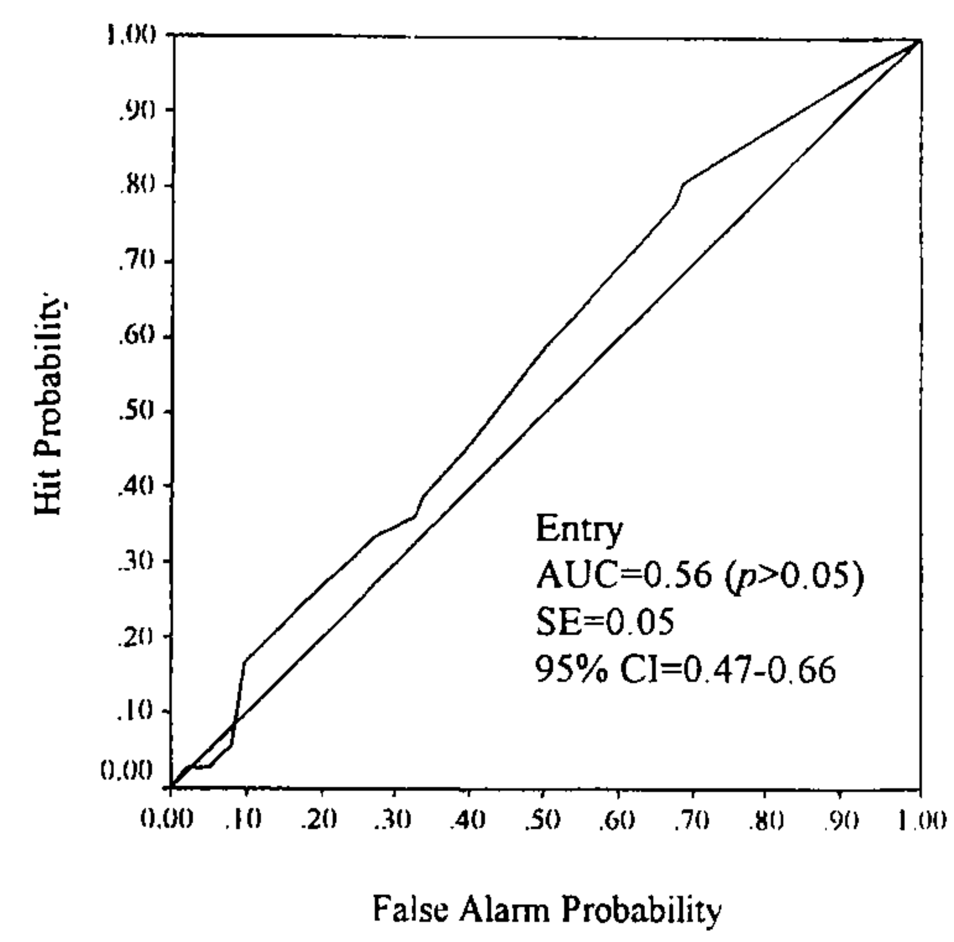
APPENDIX J
ROC RESULTS FOR MERSEYSIDE BURGLARY

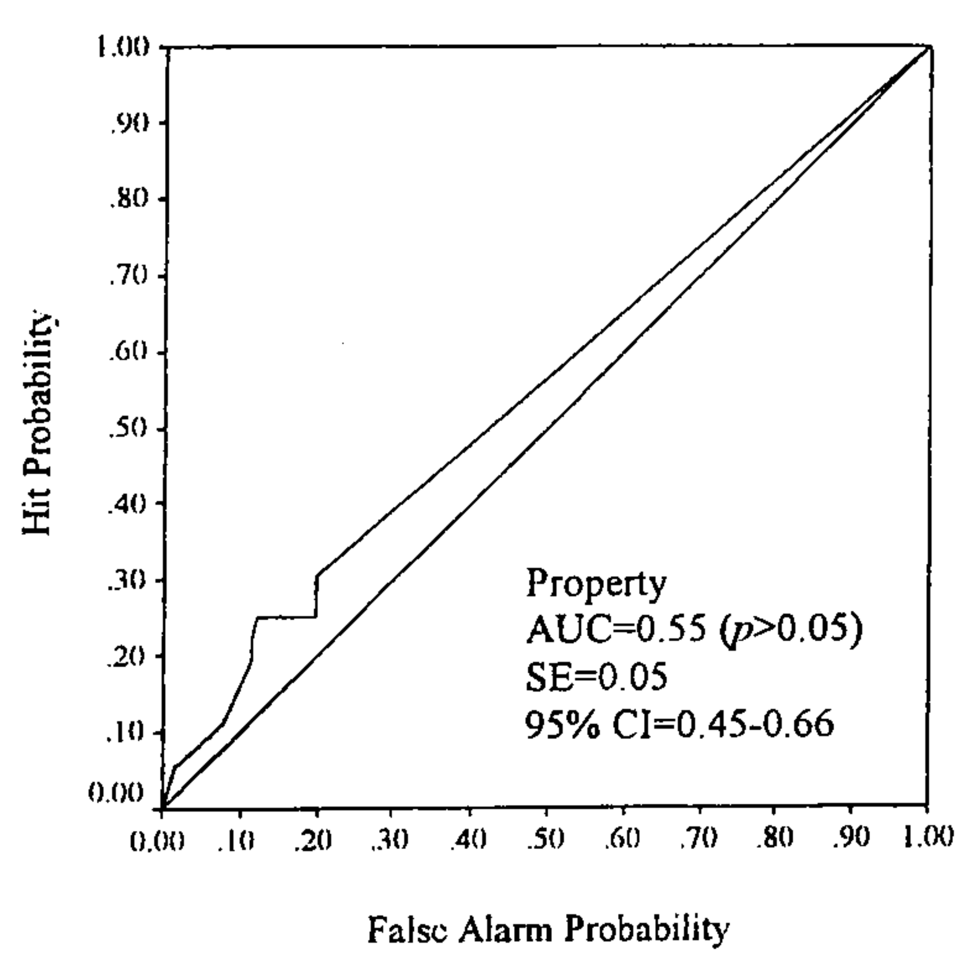
Figure J1. ROC graphs for Merseyside residential A

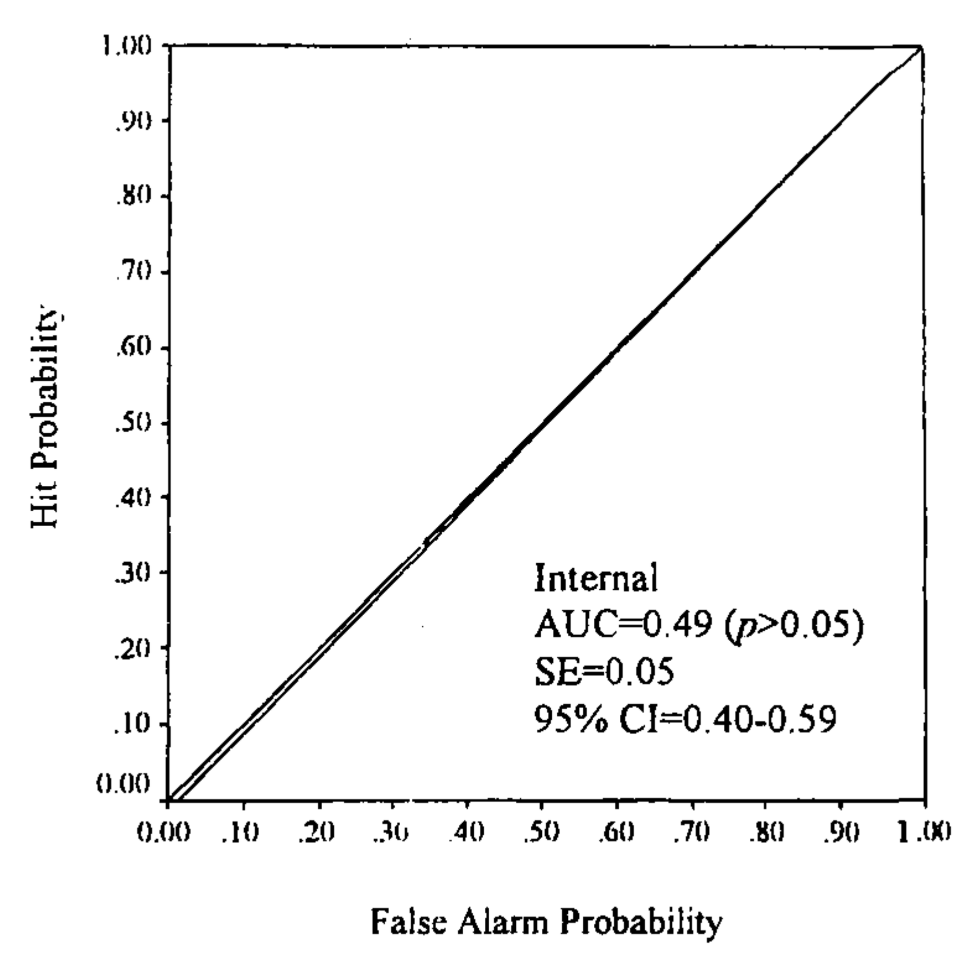












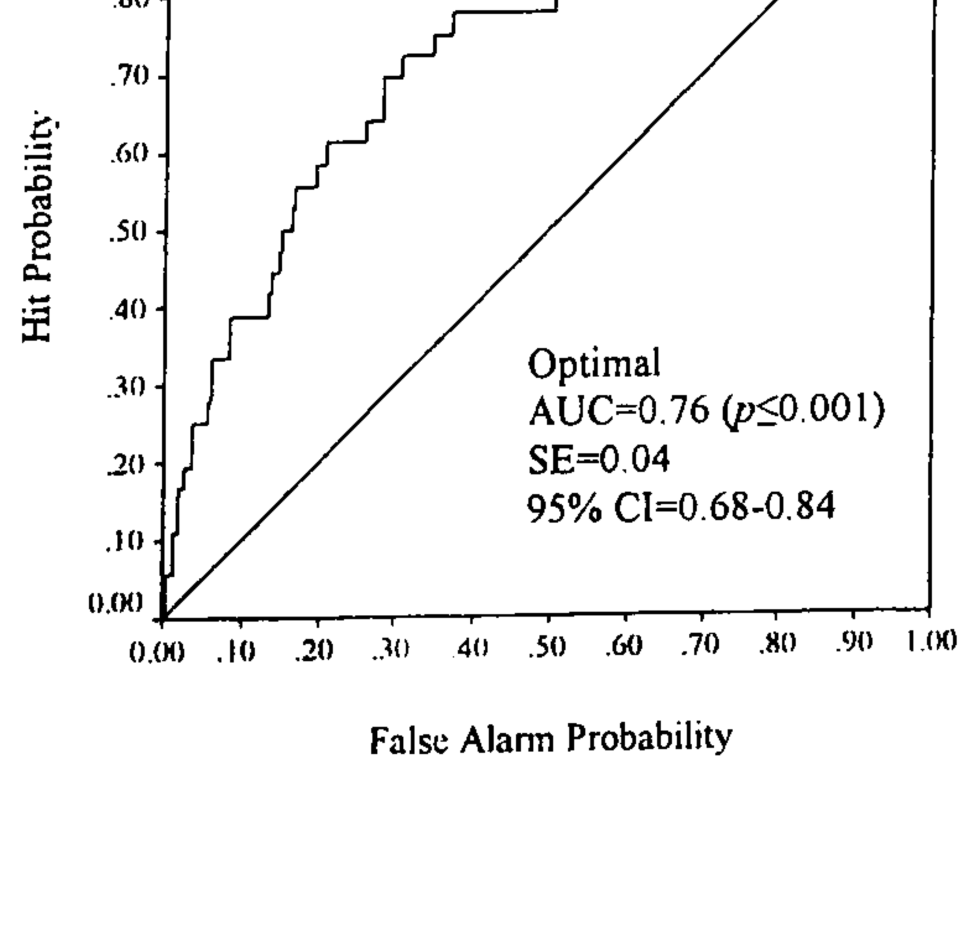


Table J1. Validation trials for Merseyside residential A

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.12$ (≤ 1.90 km)	100	0.89 (24)	0.11 (1)	0.67 (61)	0.33 (30)	10.87 (1)***
	$p \geq 0.12$ (≤ 1.90 km)	500	0.78 (39)	0.22 (11)	0.64 (288)	0.36 (162)	33.02 (1)***
2	$p \geq 0.12$ (≤ 1.90 km)	100	0.86 (6)	0.14 (1)	0.63 (59)	0.37 (34)	6.55 (1)**
	$p \geq 0.12$ (≤ 1.90 km)	500	0.76 (39)	0.24 (12)	0.65 (290)	0.35 (159)	32.28 (1)***
3	$p \geq 0.12$ (≤ 1.90 km)	100	0.67 (8)	0.33 (4)	0.73 (64)	0.27 (24)	7.53 (1)**
	$p \geq 0.12$ (≤ 1.90 km)	500	0.75 (38)	0.25 (13)	0.65 (292)	0.35 (157)	30.10 (1)***
4	$p \geq 0.12$ (≤ 1.90 km)	100	0.69 (9)	0.31 (4)	0.68 (59)	0.32 (28)	6.66 (1)**
	$p \geq 0.12$ (≤ 1.90 km)	500	0.71 (40)	0.29 (16)	0.67 (296)	0.33 (148)	30.76 (1)***
5	$p \geq 0.12$ (≤ 1.90 km)	100	0.80 (8)	0.20 (2)	0.77 (69)	0.23 (21)	14.04 (1)***
	$p \geq 0.12$ (≤ 1.90 km)	500	0.75 (33)	0.25 (11)	0.66 (299)	0.34 (157)	28.03 (1)***
Average	$p \geq 0.12$ (≤ 1.90 km)	100	0.78 (7.80)	0.22 (2.40)	0.70 (62.40)	0.30 (27.40)	--
	$p \geq 0.12$ (≤ 1.90 km)	500	0.75 (37.80)	0.25 (12.60)	0.65 (293.00)	0.35 (156.60)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Figure J2. ROC graphs for Merseyside residential B

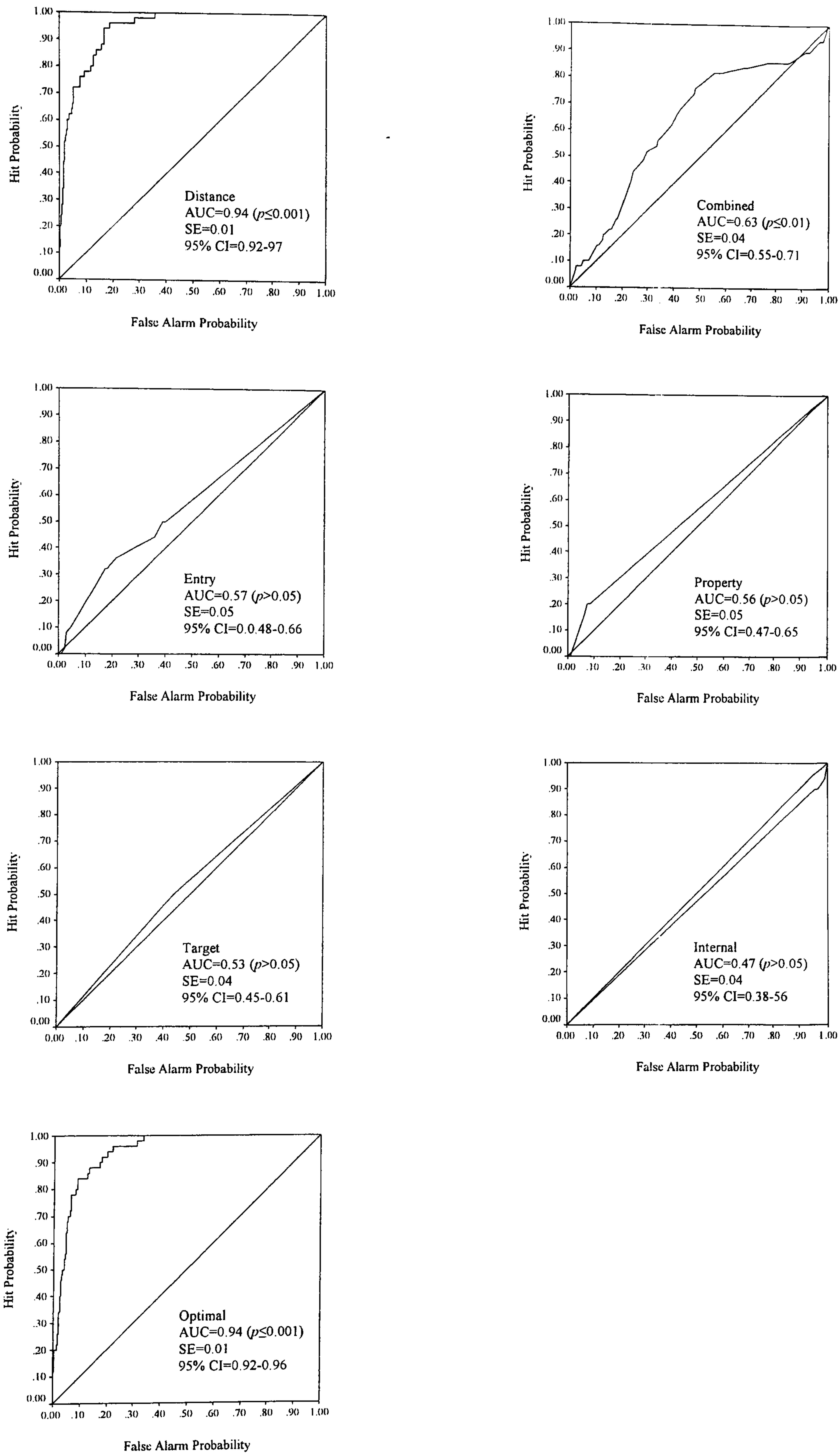


Table J2. Validation trials for Merseyside residential B

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.15$ (≤ 2.60 km)	500	0.76 (27)	0.24 (11)	0.85 (392)	0.15 (70)	70.17 (1)***
	$p \geq 0.15$ (≤ 2.60 km)	1000	0.84 (63)	0.16 (12)	0.86 (797)	0.14 (128)	221.02 (1)***
2	$p \geq 0.15$ (≤ 2.60 km)	500	0.88 (36)	0.12 (5)	0.87 (401)	0.13 (58)	139.31 (1)***
	$p \geq 0.15$ (≤ 2.60 km)	1000	0.76 (58)	0.24 (18)	0.86 (791)	0.14 (133)	174.26 (1)***
3	$p \geq 0.15$ (≤ 2.60 km)	500	0.69 (22)	0.31 (10)	0.84 (395)	0.16 (73)	54.98 (1)***
	$p \geq 0.15$ (≤ 2.60 km)	1000	0.79 (63)	0.21 (17)	0.86 (789)	0.14 (131)	195.89 (1)***
4	$p \geq 0.15$ (≤ 2.60 km)	500	0.78 (28)	0.22 (8)	0.87 (402)	0.13 (62)	93.92 (1)***
	$p \geq 0.15$ (≤ 2.60 km)	1000	0.74 (55)	0.26 (19)	0.85 (789)	0.15 (137)	156.53 (1)***
5	$p \geq 0.15$ (≤ 2.60 km)	500	0.92 (35)	0.08 (3)	0.84 (387)	0.16 (75)	117.79 (1)***
	$p \geq 0.15$ (≤ 2.60 km)	1000	0.79 (59)	0.21 (16)	0.86 (791)	0.14 (134)	183.47 (1)***
Average	$p \geq 0.15$ (≤ 2.60 km)	500	0.81 (29.60)	0.19 (7.40)	0.85 (395.40)	0.15 (67.60)	--
	$p \geq 0.15$ (≤ 2.60 km)	1000	0.78 (59.60)	0.22 (16.40)	0.86 (791.40)	0.14 (132.60)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Figure J3. ROC graphs for Merseyside residential C

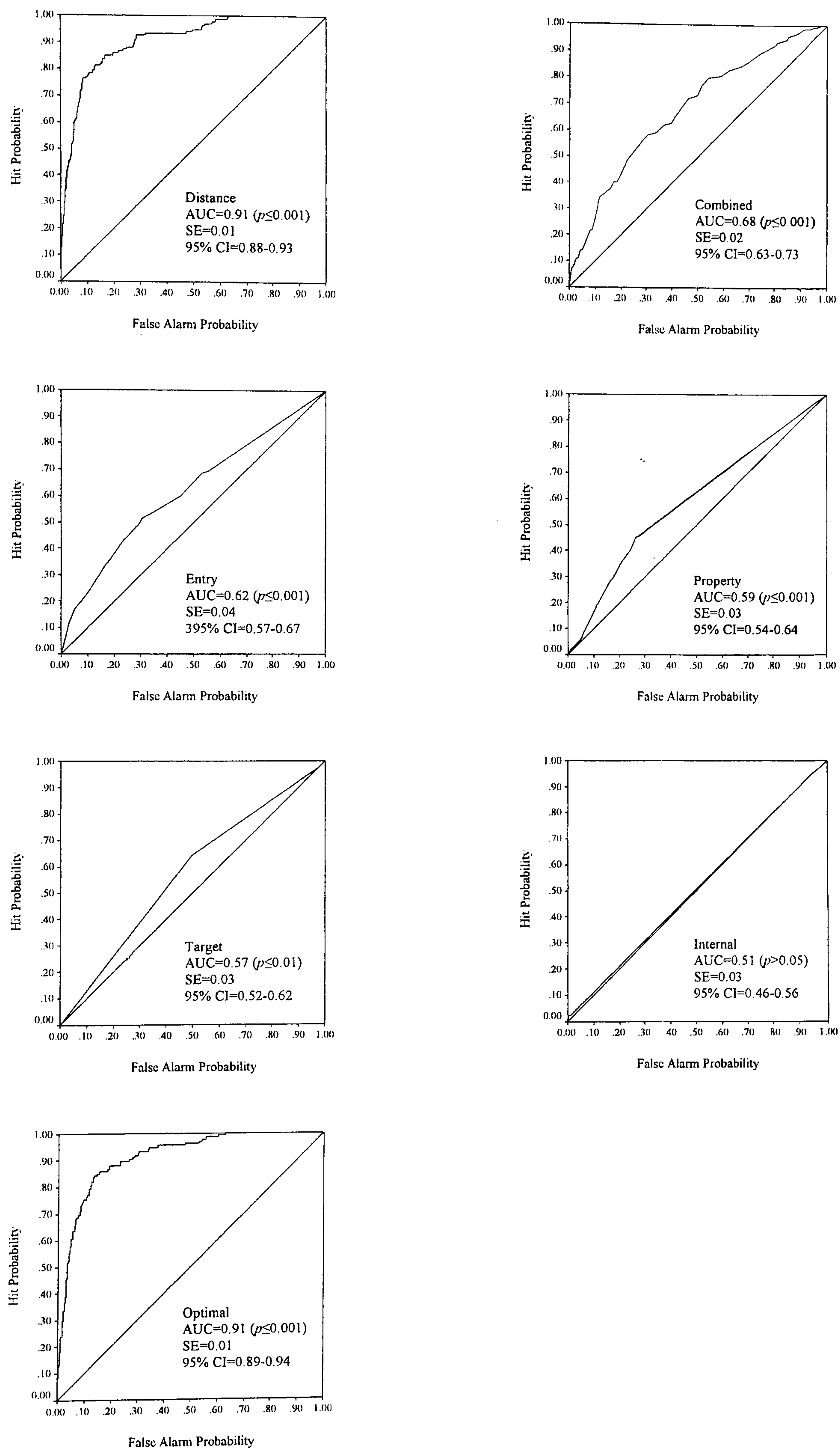


Table J3. Validation trials for Merseyside residential C

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.04$ ($\leq$2.10 km)	1000	0.79 (22)	0.21 (6)	0.83 (806)	0.17 (166)	67.42 (1)***
	$p \geq 0.04$ ($\leq$2.10 km)	5000	0.82 (124)	0.18 (27)	0.83 (4022)	0.17 (827)	402.49 (1)***
2	$p \geq 0.04$ ($\leq$2.10 km)	1000	0.79 (22)	0.21 (6)	0.83 (810)	0.17 (162)	69.41 (1)***
	$p \geq 0.04$ ($\leq$2.10 km)	5000	0.77 (105)	0.23 (32)	0.83 (4031)	0.17 (832)	310.12 (1)***
3	$p \geq 0.04$ ($\leq$2.10 km)	1000	0.78 (14)	0.22 (4)	0.87 (852)	0.13 (130)	59.73 (1)***
	$p \geq 0.04$ ($\leq$2.10 km)	5000	0.78 (108)	0.22 (30)	0.83 (4020)	0.17 (842)	323.84 (1)***
4	$p \geq 0.04$ ($\leq$2.10 km)	1000	0.78 (18)	0.22 (5)	0.82 (805)	0.18 (172)	53.72 (1)***
	$p \geq 0.04$ ($\leq$2.10 km)	5000	0.84 (119)	0.16 (23)	0.84 (4063)	0.16 (795)	420.03 (1)***
5	$p \geq 0.04$ ($\leq$2.10 km)	1000	0.81 (17)	0.19 (4)	0.84 (822)	0.16 (157)	60.28 (1)***
	$p \geq 0.04$ ($\leq$2.10 km)	5000	0.78 (116)	0.22 (32)	0.84 (4067)	0.16 (785)	376.12 (1)***
Average	$p \geq 0.04$ ($\leq$2.10 km)	1000	0.79 (18.60)	0.21 (5.00)	0.84 (819.00)	0.16 (157.40)	--
	$p \geq 0.04$ ($\leq$2.10 km)	5000	0.80 (114.40)	0.20 (28.80)	0.83 (4040.60)	0.17 (816.20)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Figure J4. ROC graphs for Merseyside residential D

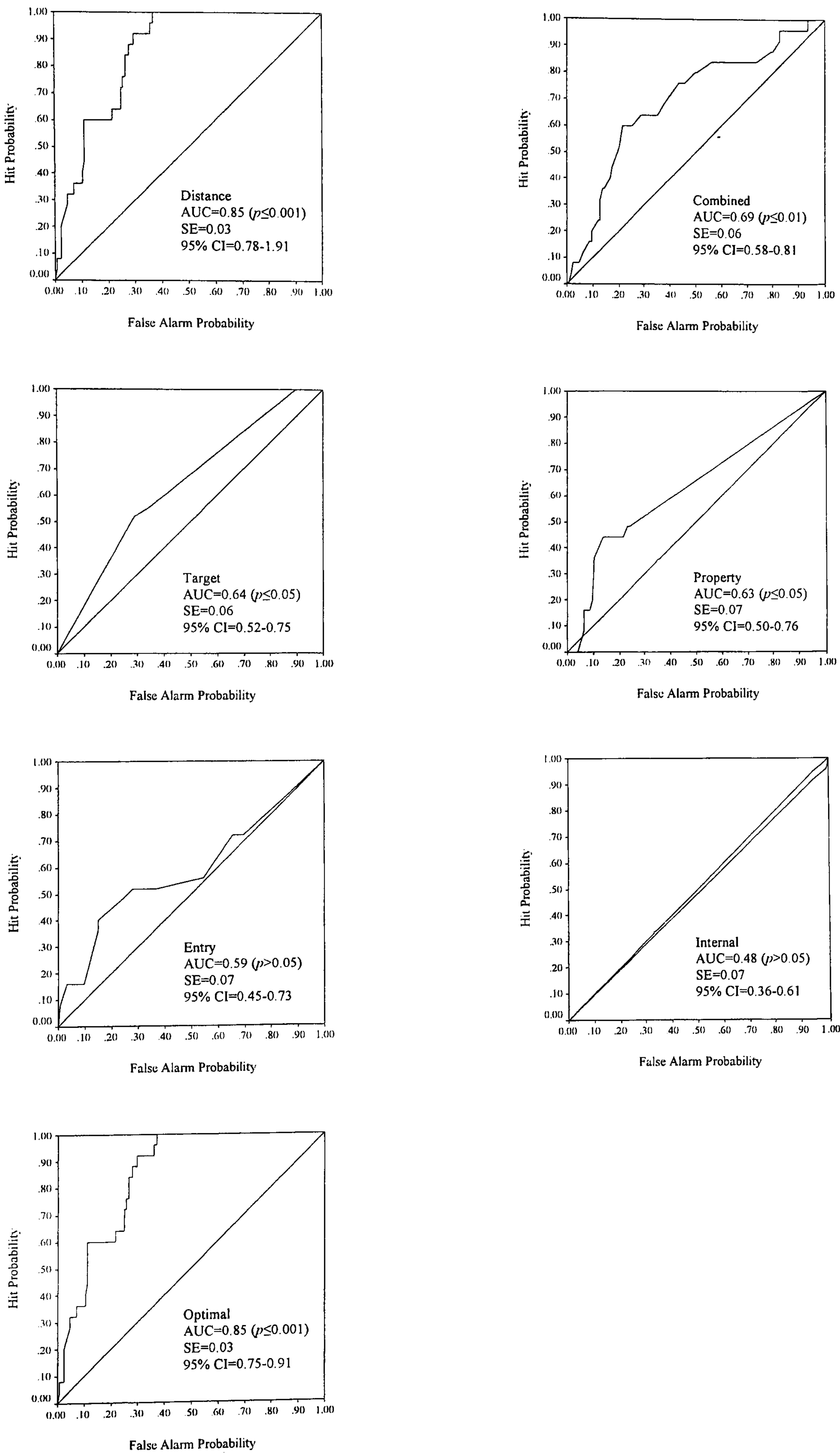


Table J4. Validation trials for Merseyside residential D

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.24$ (≤ 2.80 km)	100	0.85 (17)	0.15 (3)	0.78 (62)	0.22 (18)	27.47 (1)***
	$p \geq 0.24$ (≤ 2.80 km)	200	0.78 (29)	0.22 (8)	0.72 (117)	0.28 (46)	32.37 (1)***
2	$p \geq 0.24$ (≤ 2.80 km)	100	0.90 (18)	0.10 (2)	0.73 (58)	0.27 (22)	26.04 (1)***
	$p \geq 0.24$ (≤ 2.80 km)	200	0.73 (27)	0.27 (10)	0.74 (121)	0.26 (42)	29.74 (1)***
3	$p \geq 0.24$ (≤ 2.80 km)	100	0.82 (14)	0.18 (3)	0.72 (60)	0.28 (23)	18.07 (1)***
	$p \geq 0.24$ (≤ 2.80 km)	200	0.83 (29)	0.17 (6)	0.71 (117)	0.29 (48)	35.25 (1)***
4	$p \geq 0.24$ (≤ 2.80 km)	100	0.78 (14)	0.22 (4)	0.71 (58)	0.29 (24)	14.74 (1)***
	$p \geq 0.24$ (≤ 2.80 km)	200	0.77 (23)	0.23 (7)	0.74 (126)	0.26 (44)	29.52 (1)***
5	$p \geq 0.24$ (≤ 2.80 km)	100	0.67 (12)	0.33 (6)	0.77 (63)	0.23 (19)	13.06 (1)***
	$p \geq 0.24$ (≤ 2.80 km)	200	0.81 (25)	0.19 (6)	0.74 (125)	0.26 (44)	34.57 (1)***
Average	$p \geq 0.24$ (≤ 2.80 km)	100	0.80 (15.00)	0.20 (3.60)	0.74 (60.20)	0.26 (21.20)	--
	$p \geq 0.24$ (≤ 2.80 km)	200	0.78 (26.60)	0.22 (7.40)	0.73 (121.20)	0.27 (44.80)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Figure J5. ROC graphs for Merseyside commercial A

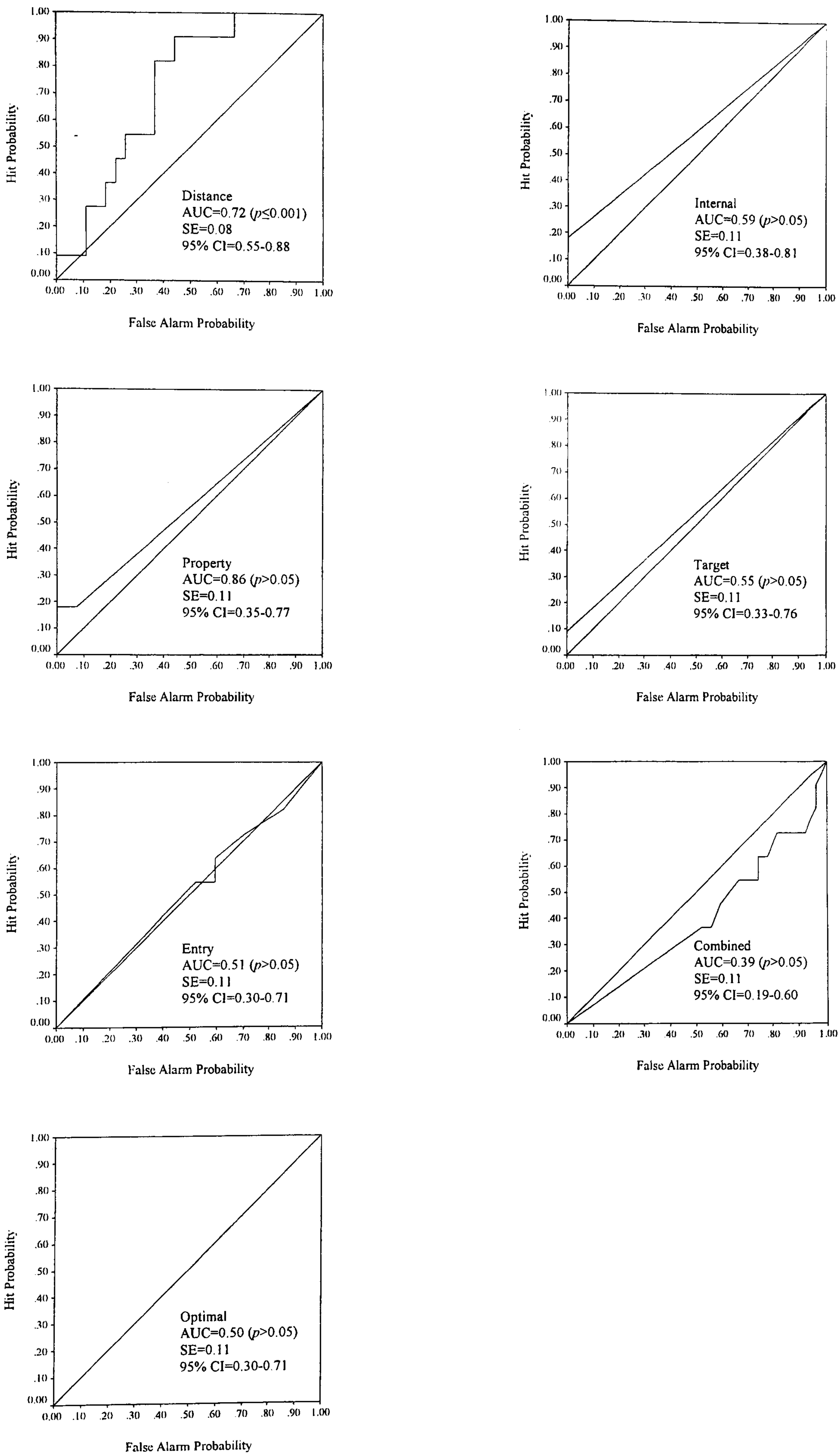


Table J5. Validation trials for Merseyside commercial A

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.30$ (≤ 5.20 km)	10	0.25 (1)	0.75 (3)	0.66 (4)	0.34 (2)	--
	$p \geq 0.30$ (≤ 5.20 km)	50	0.53 (8)	0.47 (7)	0.57 (20)	0.43 (15)	0.46 (1)
2	$p \geq 0.30$ (≤ 5.20 km)	10	1.00 (2)	0.00 (0)	0.63 (5)	0.37 (3)	--
	$p \geq 0.30$ (≤ 5.20 km)	50	0.53 (8)	0.47 (7)	0.66 (23)	0.34 (12)	1.59 (1)
3	$p \geq 0.30$ (≤ 5.20 km)	10	0.40 (2)	0.60 (3)	0.40 (2)	0.60 (3)	--
	$p \geq 0.30$ (≤ 5.20 km)	50	0.54 (7)	0.46 (6)	0.54 (20)	0.46 (17)	0.24 (1)
4	$p \geq 0.30$ (≤ 5.20 km)	10	0.75 (3)	0.25 (1)	0.33 (2)	0.67 (4)	--
	$p \geq 0.30$ (≤ 5.20 km)	50	0.50 (7)	0.50 (7)	0.58 (21)	0.42 (154)	0.28 (1)
5	$p \geq 0.30$ (≤ 5.20 km)	10	0.50 (1)	0.50 (1)	0.75 (6)	0.25 (2)	--
	$p \geq 0.30$ (≤ 5.20 km)	50	0.50 (6)	0.50 (6)	0.53 (20)	0.47 (18)	0.03 (1)
Average	$p \geq 0.30$ (≤ 5.20 km)	10	0.58 (1.80)	0.42 (1.60)	0.55 (3.80)	0.45 (2.80)	--
	$p \geq 0.30$ (≤ 5.20 km)	50	0.52 (7.20)	0.48 (6.60)	0.58 (20.80)	0.42 (15.40)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Figure J6. ROC graphs for Merseyside commercial B

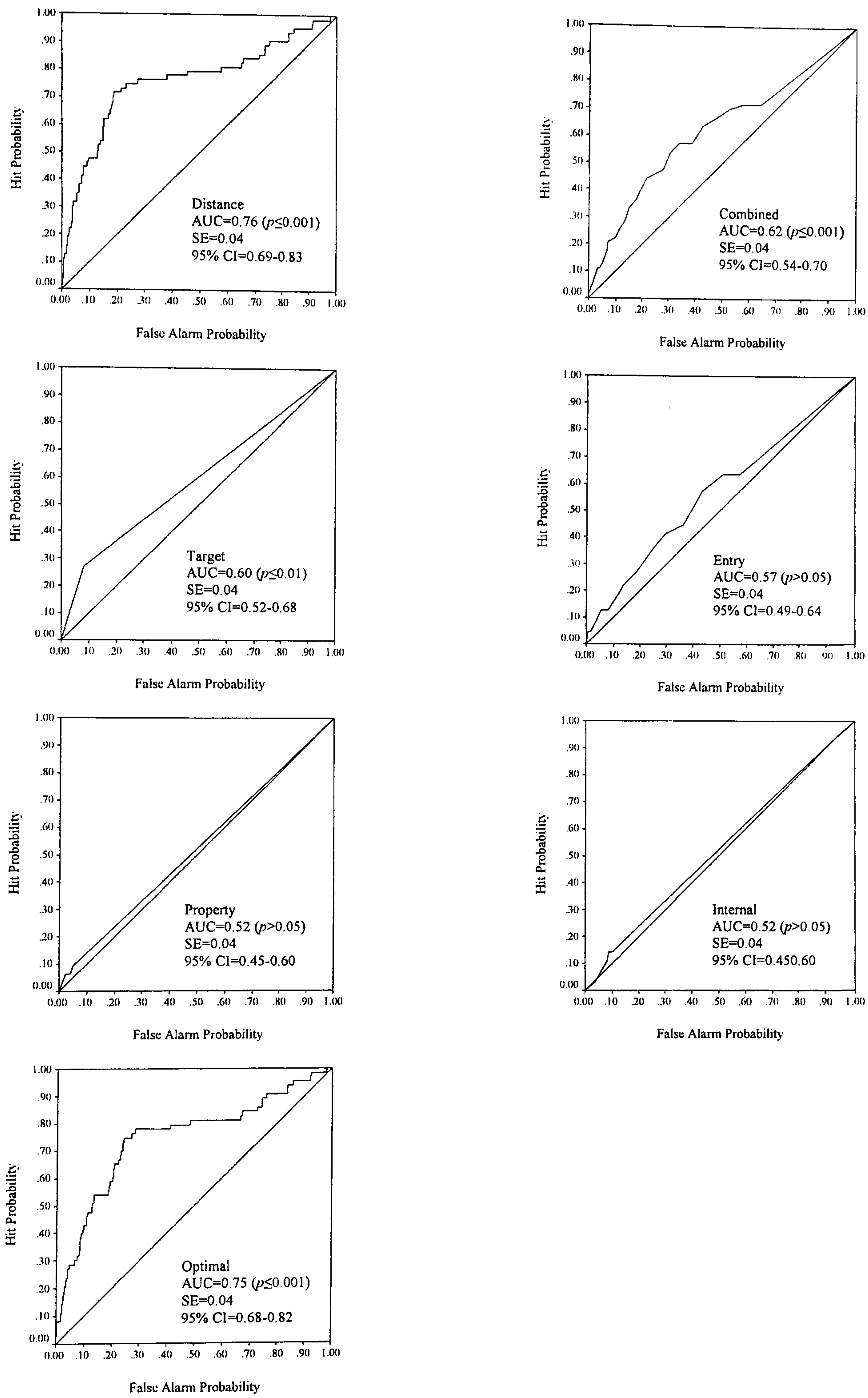


Table J6. Validation trials for Merseyside commercial B

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.01$ (≤ 3.00 km)	1000	0.77 (33)	0.23 (10)	0.65 (618)	0.35 (339)	30.08 (1)***
	$p \geq 0.01$ (≤ 3.00 km)	2000	0.78 (74)	0.22 (21)	0.62 (1177)	0.38 (728)	59.31 (1)***
2	$p \geq 0.01$ (≤ 3.00 km)	1000	0.78 (38)	0.22 (11)	0.63 (598)	0.37 (353)	36.99 (1)***
	$p \geq 0.01$ (≤ 3.00 km)	2000	0.76 (71)	0.24 (22)	0.62 (1181)	0.38 (726)	54.19 (1)***
3	$p \geq 0.01$ (≤ 3.00 km)	1000	0.75 (30)	0.25 (10)	0.62 (598)	0.38 (362)	22.41 (1)***
	$p \geq 0.01$ (≤ 3.00 km)	2000	0.75 (77)	0.25 (25)	0.63 (1190)	0.37 (708)	59.20 (1)***
4	$p \geq 0.01$ (≤ 3.00 km)	1000	0.84 (36)	0.16 (7)	0.63 (605)	0.37 (352)	38.18 (1)***
	$p \geq 0.01$ (≤ 3.00 km)	2000	0.79 (83)	0.21 (22)	0.62 (1179)	0.38 (716)	70.61 (1)***
5	$p \geq 0.01$ (≤ 3.00 km)	1000	0.78 (39)	0.22 (11)	0.63 (599)	0.37 (351)	33.65 (1)***
	$p \geq 0.01$ (≤ 3.00 km)	2000	0.76 (73)	0.24 (23)	0.63 (1205)	0.37 (699)	59.65 (1)***
Average	$p \geq 0.01$ (≤ 3.00 km)	1000	0.78 (35.20)	0.22 (9.80)	0.63 (603.60)	0.37 (351.40)	--
	$p \geq 0.01$ (≤ 3.00 km)	2000	0.77 (75.60)	0.23 (22.60)	0.62 (1186.40)	0.38 (715.40)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Figure J7. ROC graphs for Merseyside commercial C

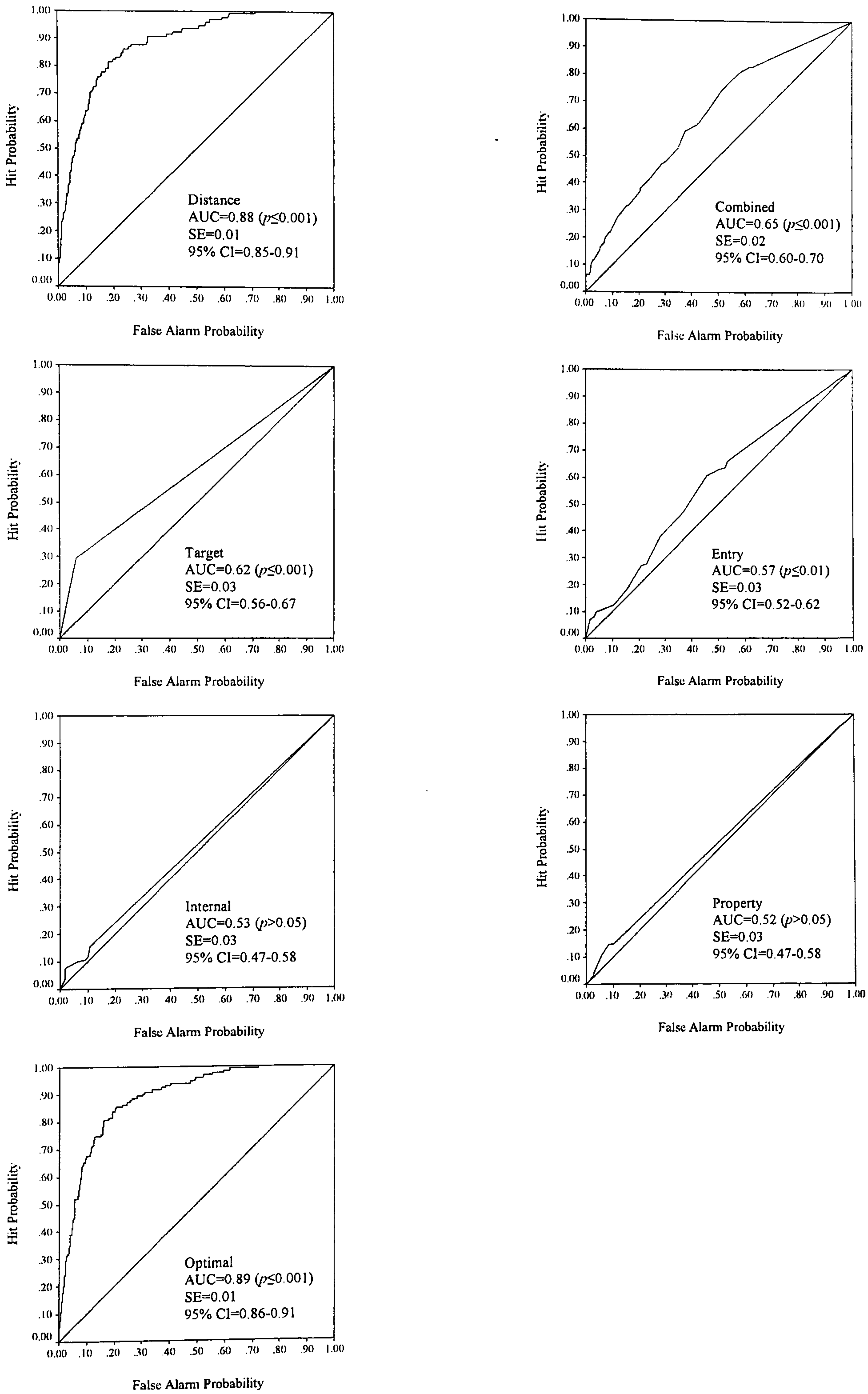


Table J7. Validation trials for Merseyside commercial C

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.01$ (≤ 2.30 km)	1000	0.96 (23)	0.04 (1)	0.72 (695)	0.28 (281)	49.76 (1)***
	$p \geq 0.01$ (≤ 2.30 km)	5000	0.86 (116)	0.14 (19)	0.71 (3432)	0.29 (1433)	195.90 (1)***
2	$p \geq 0.01$ (≤ 2.30 km)	1000	0.88 (22)	0.12 (3)	0.71 (692)	0.29 (283)	40.00 (1)***
	$p \geq 0.01$ (≤ 2.30 km)	5000	0.91 (134)	0.09 (13)	0.71 (3431)	0.29 (1422)	254.67 (1)***
3	$p \geq 0.01$ (≤ 2.30 km)	1000	0.79 (19)	0.21 (5)	0.71 (693)	0.29 (283)	27.97 (1)***
	$p \geq 0.01$ (≤ 2.30 km)	5000	0.89 (117)	0.11 (14)	0.71 (3458)	0.29 (1411)	218.83 (1)***
4	$p \geq 0.01$ (≤ 2.30 km)	1000	0.89 (24)	0.11 (3)	0.71 (686)	0.29 (287)	43.25 (1)***
	$p \geq 0.01$ (≤ 2.30 km)	5000	0.88 (115)	0.12 (16)	0.71 (3435)	0.29 (1434)	203.02 (1)***
5	$p \geq 0.01$ (≤ 2.30 km)	1000	0.88 (22)	0.12 (3)	0.73 (710)	0.27 (265)	44.06 (1)***
	$p \geq 0.01$ (≤ 2.30 km)	5000	0.88 (115)	0.12 (15)	0.71 (3439)	0.29 (1431)	206.90 (1)***
Average	$p \geq 0.01$ (≤ 2.30 km)	1000	0.88 (22.00)	0.12 (3.00)	0.72 (695.20)	0.28 (279.80)	--
	$p \geq 0.01$ (≤ 2.30 km)	5000	0.88 (119.40)	0.12 (15.40)	0.71 (3439.00)	0.29 (1426.20)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Figure J8. ROC graphs for Merseyside commercial D

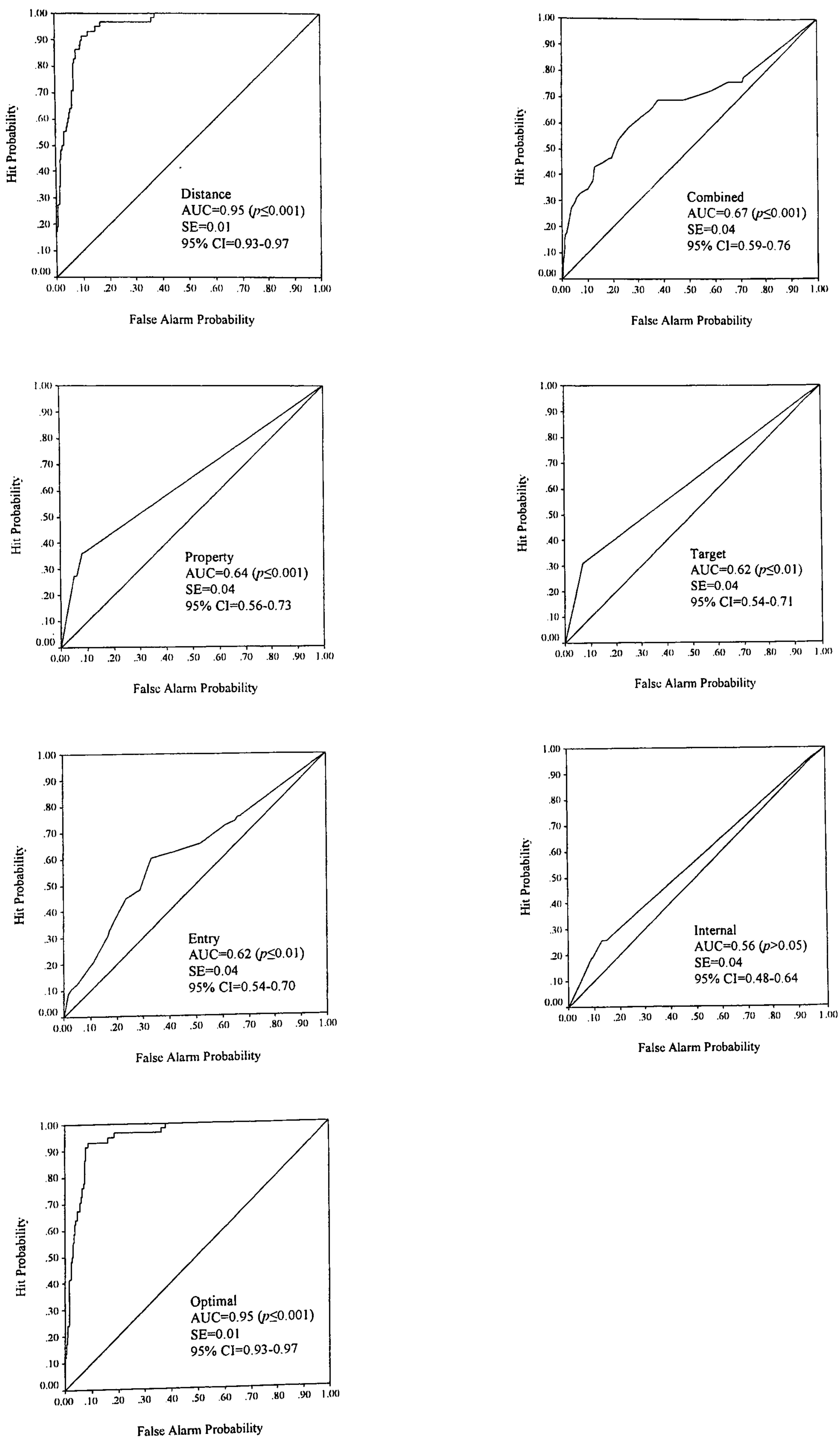


Table J8. Validation trials for Merseyside commercial D

Sample	Threshold (distance)	Sample size	*p*H (freq.)	*p*M (freq.)	*p*CR (freq.)	*p*FA (freq.)	X^2 (df)
1	$p \geq 0.17$ (≤ 2.30 km)	500	0.73 (24)	0.27 (9)	0.90 (421)	0.10 (46)	101.21 (1)***
	$p \geq 0.17$ (≤ 2.30 km)	1000	0.85 (57)	0.15 (10)	0.91 (849)	0.09 (84)	298.67 (1)***
2	$p \geq 0.17$ (≤ 2.30 km)	500	0.85 (29)	0.15 (5)	0.90 (419)	0.10 (47)	139.06(1)***
	$p \geq 0.17$ (≤ 2.30 km)	1000	0.83 (57)	0.17 (12)	0.90 (838)	0.10 (93)	265.70 (1)***
3	$p \geq 0.17$ (≤ 2.30 km)	500	0.86 (25)	0.14 (4)	0.91 (429)	0.09 (42)	140.63 (1)***
	$p \geq 0.17$ (≤ 2.30 km)	1000	0.81 (57)	0.19 (13)	0.92 (852)	0.08 (78)	297.42 (1)***
4	$p \geq 0.17$ (≤ 2.30 km)	500	0.77 (30)	0.23 (9)	0.89 (410)	0.11 (51)	114.89 (1)***
	$p \geq 0.17$ (≤ 2.30 km)	1000	0.87 (54)	0.13 (8)	0.90 (847)	0.10 (91)	280.99 (1)***
5	$p \geq 0.17$ (≤ 2.30 km)	500	0.87 (34)	0.13 (5)	0.93 (427)	0.07 (34)	194.89 (1)***
	$p \geq 0.17$ (≤ 2.30 km)	1000	0.81 (50)	0.19 (12)	0.91 (856)	0.09 (82)	262.42 (1)***
Average	$p \geq 0.17$ (≤ 2.30 km)	500	0.82 (28.40)	0.18 (6.40)	0.91 (421.20)	0.09 (44.00)	--
	$p \geq 0.17$ (≤ 2.30 km)	1000	0.83 (55.00)	0.17 (11.00)	0.91 (848.40)	0.09 (85.60)	--

*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

Table J9. Optimal thresholds for Merseyside burglary data

	Merseyside residential				Merseyside commercial			
	A	B	C	D	A	B	C	D
Distance	$p \geq 0.12$; ≤ 1.90	$p \geq 0.15$; 2.60	$p \geq 0.04$; ≤ 2.10	$p \geq 0.24$; ≤ 2.20	$p \geq 0.30$; ≤ 5.20	$p \geq 0.01$; ≤ 3.00	$p \geq 0.01$; ≤ 2.30	$p \geq 0.17$; ≤ 2.30
Combined	$p \geq 0.08$; ≥ 0.23	$p \geq 0.07$; ≥ 0.18	$p \geq 0.02$; ≥ 0.20	$p \geq 0.17$; ≥ 0.23	$p \geq 0.30$; --*	$p \geq 0.04$; ≥ 0.08	$p \geq 0.02$; ≥ 0.05	$p \geq 0.06$; ≥ 0.17
Target	$p \geq 0.11$; ≥ 0.80	$p \geq 0.08$; ≥ 0.70	$p \geq 0.03$; ≥ 0.77	$p \geq 0.14$; ≥ 0.40	$p \geq 0.18$; ≥ 0.01	$p \geq 0.04$; ≥ 0.01	$p \geq 0.03$; ≥ 0.10	$p \geq 0.08$; ≥ 0.18
Entry	$p \geq 0.09$; ≥ 0.20	$p \geq 0.06$; ≥ 0.01	$p \geq 0.02$; ≥ 0.06	$p \geq 0.15$; ≥ 0.18	$p \geq 0.34$; --*	$p \geq 0.05$; ≥ 0.21	$p \geq 0.03$; ≥ 0.23	$p \geq 0.06$; ≥ 0.18
Property	$p \geq 0.09$; ≥ 0.02	$p \geq 0.04$; ≥ 0.01	$p \geq 0.03$; ≥ 0.12	$p \geq 0.10$; ≥ 0.01	$p \geq 0.30$; ≥ 0.02	$p \geq 0.03$; ≥ 0.01	$p \geq 0.01$; ≥ 0.01	$p \geq 0.04$; ≥ 0.01
Internal	$p \geq 0.07$; ≥ 0.01	$p \geq 0.09$; --*	$p \geq 0.02$; ≥ 0.01	$p \geq 0.50$; --*	$p \geq 0.32$; ≥ 0.03	$p \geq 0.03$; ≥ 0.01	$p \geq 0.01$; ≥ 0.01	$p \geq 0.04$; ≥ 0.01
Optimal	$p \geq 0.12$	$p \geq 012$	$p \geq 0.05$	$p \geq 0.24$	$p \geq 0.50$	$p \geq 0.06$	$p \geq 0.04$	$p \geq 0.19$

Tosc: Instrument brought to scene
Frsc: Instrument used from scene
Bdyfrc: Bodily force
Frclock: Forced lock