



Citation for published version:

Slater, C, Woodhams, J & Hamilton-giachritsis, C 2015, 'Testing the assumptions of crime linkage with stranger sex offenses: a more ecologically-valid study', *Journal of Police and Criminal Psychology*, vol. 30, no. 4, pp. 261-273. <https://doi.org/10.1007/s11896-014-9160-3>

DOI:

[10.1007/s11896-014-9160-3](https://doi.org/10.1007/s11896-014-9160-3)

Publication date:

2015

Document Version

Early version, also known as pre-print

[Link to publication](#)

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Testing the Assumptions of Crime Linkage with Stranger Sex Offenses: A More Ecologically-Valid
Study

Chelsea Slater, Jessica Woodhams and Catherine Hamilton-Giachritsis
University of Birmingham

To cite this article:

Slater, C., Woodhams, J., & [Hamilton-Giachritsis](#), C.E. (2014). Testing the assumptions of crime linkage with stranger sex offenses: A more ecologically valid study. *Journal of Police and Criminal Psychology*. DOI: 10.1007/s11896-014-9160-3.

Link to this article:

<http://link.springer.com/article/10.1007/s11896-014-9160-3/fulltext.html>

We would like to acknowledge the assistance of the Serious Crime Analysis Section (National Crime Agency, UK) for allowing the sampling of the data in their care for this research project.

Author Note

Chelsea Slater, School of Psychology, University of Birmingham, Frankland Building, Edgbaston, Birmingham B15 2TT, UK; Jessica Woodhams, School of Psychology, University of Birmingham, Frankland Building, Edgbaston, Birmingham B15 2TT, UK; Catherine Hamilton-Giachritsis, Department of Psychology, University of Bath, Claverton Down, Bath BA2 7AY, UK

Correspondence concerning this article should be addressed to Chelsea Slater, School of Psychology, Frankland Building, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK. Email: CLS011@bham.ac.uk.

Abstract

An increasing amount of research has been conducted on crime linkage, a practice that has already been presented as expert evidence in some countries; however it is questionable whether standards of admissibility, applied in some jurisdictions, have been achieved (e.g., the Daubert criteria). Much research has assessed the two basic assumptions underpinning this practice: that offenders are consistent in the way they commit their crimes and that offenders commit their crimes in a relatively distinctive manner. While studies of these assumptions with stranger sex offenses exist, they are problematic for two reasons: (1) small samples (usually < 30 series), and (2) samples consisting solely of serial offenses. The current study improved on past research through the use of a much larger dataset ($N=50$ series, 194 offenses; and $N= 50$ one-off offenses) and by sampling the offenses of both serial and one-off sex offenders, thereby representing a more ecologically valid test of the assumptions. The two assumptions were tested simultaneously by assessing how accurately 365 linked crime pairs could be differentiated from 29,281 unlinked crime pairs through the use of Leave-One-Out Cross-Validation logistic regression followed by Receiver Operating Characteristic analysis. An excellent level of predictive accuracy was achieved providing support for the assumptions underpinning crime linkage.

Keywords: linkage analysis, rapists, sex offenders, comparative case analysis, behavioral consistency

Testing the Assumptions of Crime Linkage with Stranger Sex Offenses: A More Ecologically-Valid Study

1. Introduction

Crime linkage is a police operational practice whereby police records are analyzed with the aim of identifying similarities in behavior between two or more crimes that would suggest they were committed by the same offender (Woodhams & Grant, 2006). It is an area of research that has grown significantly in the past 10 years and has included investigations of sexual assaults and rape (e.g., Grubin, et al., 2001), homicides (e.g., Santtila et al., 2008), arson (e.g., Ellingwood et al., 2013), burglaries (e.g., Bennell & Jones, 2005), robberies (e.g., Lin & Brown, 2006), and car theft (e.g., Tonkin, et al., 2008). Notably, research has focused on the application of crime linkage with ‘hard to solve’ crimes such as those committed by a stranger who had no previous connection to the victim (e.g., Santtila, et al., 2005).

An impetus for research in this field has been the use of crime linkage analysis to advise police investigations, as well as its use in legal proceedings in some countries (Bosco, et al., 2010; Charron & Woodhams, 2010; Hazelwood & Warren, 2004; Labuschagne, 2006), necessitating a comprehensive test of its two underlying principles. First, this paper outlines the ways in which crime linkage is utilized in practice and what this means for testing the two underlying assumptions. It then briefly describes the existing studies of these assumptions with stranger sexual offenses, explaining their limitations and how these affect the conclusions that can be drawn. The remainder of the paper presents a study designed to overcome a number of key limitations, which more closely reflects the type of data with which crime linkage would be conducted in reality. It therefore represents a more ecologically valid test of the underlying principles of crime linkage.

1.1. Crime Linkage in Practice

Although crime linkage is a relatively new area of research in legal and criminological psychology, with most research having been conducted in the last decade, it has been practiced for far longer than this. For example, the Violent Criminal Apprehension Program (ViCAP) of the Federal Bureau of Investigation (FBI) used to support crime linkage was devised in 1985 (Royal Canadian

Mounted Police [RCMP], n.d.). It is also widely practiced across several countries including the United Kingdom, the United States, Australia, South Africa, Belgium, the Czech Republic, France, Germany, Ireland, the Netherlands, New Zealand, and Switzerland (Bosco et al., 2010; Hazelwood & Warren, 2004; Labuschagne, 2006; RCMP, n.d.).

The widespread use of crime linkage can be explained by the advantages it holds for police forces if conducted accurately. For example, the processing of physical evidence can take weeks or months in some countries and while waiting for such evidence, crime linkage can be used to identify potential linked cases that can be investigated while awaiting results regarding physical evidence (Labuschagne, 2012). These might be cases where the suspect also remains unknown. In such a scenario, the identification of these “linked” crimes could facilitate the pooling of other forensic evidence from each crime scene (Davies, 1991). Alternatively, for one or more “linked” crimes a suspect could have been identified thereby expediting the apprehension of a suspect for the entire series. Davies (1991) also argues that behavioral linking can enhance the credibility of victims because each victim gains credibility from the others. This has been demonstrated by Jordan (2001, as cited by Kelly, 2010) whereby cases first deemed as false allegations by the police were subsequently perceived to be credible when another similar offense occurred.

The use of crime linkage is not limited to advising police investigations; it has also been used to prosecute an individual for multiple crimes. Evidence of behavioral similarity and distinctiveness has been introduced into legal proceedings in the UK, the US, Australia and South Africa (Bosco et al., 2010; Charron & Woodhams, 2010; Hazelwood & Warren, 2004; Labuschagne, 2006; Meyer, 2007; Woodhams & Toye, 2007) to suggest that the same offender was responsible (or not) for two or more crimes (Bosco et al., 2010). Typically, at least one other form of evidence is also presented (e.g., eyewitness identification, confession) that links the offender to one or more of the crimes he/she is being tried for, but this evidence is absent or weak for other offenses (see Labuschagne, 2006, for a case example).

The admittance of such evidence into legal proceedings is governed in some countries by particular rules. In the US, the admissibility of scientific evidence into the courts is controlled by a

conglomeration of court cases and federal rules (Meyer, 2007). Often it is the decision of the individual state which precedent to follow. The generally accepted standard for many states is the Daubert criteria (1993), whereas others follow the older Frye standard (1923) or the more recent *Federal Rules of Evidence*, specifically Rule 702 (Groscup et al., 2002).

1.1.1. The Frye standard. This standard was the precursor to the Daubert criteria and Rule 702. It requires that the method by which the evidence was obtained is accepted by experts in that field.

1.1.2. The Daubert criteria. This has five main points that must be met in order for the evidence to be presented in the courts;

1. The ability for empirical testing of the theory or technique (and that it has been tested);
2. The evidence must have been subjected to peer review and publication;
3. There needs to be a known or potential error rate for the practice;
4. There must be standards and guidelines for the practice;
5. The degree to which the theory or technique is generally accepted by the scientific community.

In 2007, Woodhams, Bull and Hollin considered how well the field of crime linkage fared against each of the Daubert criteria. With respect to criterion 1, they explained that crime linkage is underpinned by two assumptions (or theories); the Assumption of Behavioral Consistency, which states that offenders are consistent in the way they commit their crimes (Canter, 1995), and the Assumption of Behavioral Distinctiveness, which states that offenders will commit their crimes in a relatively distinctive manner (Bennell & Canter, 2002). The study of behavioral consistency and distinctiveness has a long history in personality psychology (e.g. Allport, 1937) where models (e.g., the Cognitive Affective Personality System, Mischel & Shoda, 1995) have been proposed that suggest behavioral consistency will be observed in situations of psychological similarity and that distinctiveness in individuals' behavior displayed in the same situation emerges because of differences in learning histories, personality dispositions and so on.

In terms of criterion 1, Woodhams et al. (2007a) concluded that while these two assumptions could be tested and that some efforts had been made to test them, there were some significant limitations with the research, including a lack of realism (see below for a fuller discussion). They further described how crime linkage had been subject to peer review and publication (criterion 2) but that there was not a known error rate (criterion 3). With regards to the last two criteria (4 and 5), whether crime linkage theories or techniques are generally accepted by the scientific field and if there are standards and guidelines, Woodhams et al. concluded “not yet”. However, in the six years since Woodhams, et al. (2007a) presented this critique, several studies have been published, specifically on sex offending and crime linkage, which are discussed in more detail below.

1.1.3. Rule 702. The requirements of Rule 702 (*Federal Rules of Evidence, amended 2011*) are broader:

“A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if: (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case.”

Of most relevance to the research of crime linkage is point (c). This rule requires expert testimony utilizing crime linkage analysis to be a product of reliable principles *and* methods. Specifically, the current paper focuses on the state of the research with regards to (c)(i), whether its principles (i.e., behavioral consistency and behavioral distinctiveness) are reliable. Other researchers have investigated (c)(ii), the methods, and interested parties are referred to Snook, Luther, House, Bennell and Taylor (2012).

In summary, the field of crime linkage must ensure that research meets the standards required for admission in court. Therefore, this paper aims to expand knowledge on the reliability of the underlying principles of crime linkage for serial sex offenses.

1.2. Serial Sex Offending and Tests of the Crime Linkage Assumptions

Given the use of crime linkage analysis in legal proceedings, and that rules of evidence require its principles to have been tested (and supported), it is unsurprising that the main focus of crime linkage research has been to test their validity. Serial sex offenses are one of the crime types most often the subject of these empirical tests. Studies have centered on sexual offenses committed by offenders who are strangers to their victims because these cases are often more difficult to solve than offenses committed by a victim's acquaintance (Canter, 1996; Roberts & Grossman, 1993) and it is on such cases that crime linkage would be conducted in practice.

The two basic assumptions have been studied in a variety of ways within the research on linking sexual offenses. Some studies have assessed only the degree of behavioral consistency exhibited by serial sex offenders whereas others have assessed both assumptions simultaneously using a variety of statistical methods, including discriminant function analysis and multidimensional scaling (e.g., Santtila et al., 2005), and logistic regression and ROC analysis (e.g., Bennell, et al., 2009).

In the consistency studies, behavioral consistency has been quantified using Jaccard's coefficient, percentage agreement and kappa, with the former being most common. Sjöstedt, et al. (2004) found serial sex offenders, particularly those who had targeted stranger victims, to show some stability in their victim selection behaviors¹ ($\kappa > .40$). Grubin et al. (2001) assessed whether the serial sex offenders in their sample were consistent in behavioral domain types² (types of control behaviors, sex behaviors, escape behaviors and style behaviors established from cluster analysis) across their series. Eighty-three percent of the offenders were consistent in at least one domain and 26% were consistent across all four domains.

Where Jaccard's coefficient has been used to quantify the amount of similarity in crime scene behavior between two offenses, a larger coefficient indicates greater similarity in behavior. This translates

¹ Victim selection behaviors include characteristics of the victim, such as vulnerabilities or physical appearance.

² Control behaviors are those used to maintain control over the victim to facilitate the commission of the offense. Escape behaviors refer to those that enable the offender to quit the scene undetected. Style behaviors are those that do not assist in the commission of the offense but which reflect the offender's internal psychology (Grubin et al., 2001).

to greater consistency in the case of two crimes by the same offender. Jaccard's coefficients range from 0 to 1, with 0 representing absolute inconsistency in behavior and 1 representing perfect consistency. Once a Jaccard's coefficient has been calculated for all pairs in a dataset, researchers have tested the underlying assumptions of consistency and distinctiveness by contrasting the Jaccard's coefficients for linked crime pairs versus unlinked crime pairs. Linked crime pairs represent two crimes by the same serial offender whereas unlinked crime pairs are created by pairing two crimes committed by different (usually serial) offenders. If the assumptions of crime linkage are valid, the Jaccard's coefficients for the linked crime pairs should be larger (indicating greater behavioral similarity) than those of the unlinked crime pairs. For example, Mokros and Alison (2002) found the average Jaccard's coefficient for the linked crimes to be higher ($M=0.41$) than that of the unlinked pairs ($M=0.27$), suggesting support for the assumptions of crime linkage, however they did not determine if this difference was statistically significant. Other studies have used tests of difference to contrast linked and unlinked crime pairs. In these studies the average Jaccard's coefficient for the linked pairs has ranged from .39 to .52 but only .17 to .34 for the unlinked pairs (Bennell et al., 2009; Bennell et al., 2010; Woodhams, et al., 2007b; Woodhams & Labuschagne, 2012). In all cases, the average Jaccard's coefficient for the linked pairs was significantly larger indicating greater behavioral similarity.

The studies that have utilized more advanced statistical tests to assess the underlying principles of crime linkage with sex offenses tend to fall into one of two categories: 1) statistical tests which assess the accuracy with which a query crime can be allocated to the correct series or with which crimes from the same series as the 'query crime' can be identified (e.g., Grubin et al., 2001; Santtila et al., 2005); or 2) statistical tests which determine how accurately linked crime pairs can be distinguished from unlinked crime pairs based on behavioral similarity (e.g., Bennell et al., 2009; Woodhams & Labuschagne, 2012). For the first approach, accuracy rate(s) are compared to what would be expected by chance alone to see if they differ. As Tonkin et al. point out in their 2012 study, comparing the statistical model against chance might not be a fair comparison, as practitioners may perform at a level better (or worse) than chance. For the second approach, Receiver Operating Characteristic (ROC) analysis is used to quantify discrimination

accuracy with the area under the curve (AUC) indicating how accurately linked and unlinked crime pairs can be distinguished. AUC values range from 0 to 1.00 with a larger value representing greater accuracy.

With regards to the first approach, several studies have extracted a query crime from their research database and then rank ordered all other remaining offenses according to their behavioral similarity to the query offense (Grubin et al., 2001; Santtila et al., 2005). This analysis has been conducted using bespoke computer algorithms (Grubin et al., 2001), multidimensional scaling, and discriminant function analysis (Santtila et al., 2005). Having done this, a pre-specified top percentage of ranked crimes are examined to determine how many are actually from the same series as the query offense and whether this is more than would be expected by chance. For all but two of the 117 series examined by Grubin et al. (2001) the figures were significantly higher than would be expected by chance. Santtila et al. (2005) reported that another crime from the same series was within the top five most similar crimes more than 40% of the time and nearly 60% of the time within the top ten cases found by the linking model. Overall, the accuracy of offense allocation to series was significantly greater than chance (at 25.6%) and for 86% of the crimes, the correct series was within the top 10 to which it was predicted to belong. Similarly, using a bespoke computer algorithm, Yokota, et al. (2007) used behavioral similarity to predict to which *offender* each offense in their database belonged. The top five percent of offenders in each ranked list were scrutinized to determine whether the correct offender was present. Twenty-four of the 81 offenders were correctly ranked number one for their offense, and the median rank position for the correct offender was four (Range = 1-339). Several studies have therefore demonstrated that it is possible to allocate offenses to the correct series or to offender at rates significantly higher than chance. However, it should also be noted that these figures illustrate that there is still quite a considerable degree of error in making such predictions.

This approach to evaluating crime linkage principles has been criticized by Bennell and colleagues for using just one decision-threshold (e.g., specifying a particular cut-off – “top ten”), which can produce a distorted picture of linking accuracy (Bennell et al., 2009). Instead, Bennell (2005) proposed Receiver Operating Characteristic (ROC) analysis as the “gold standard” analysis since it is able

to evaluate linking accuracy across a range of thresholds (as well as having a number of other advantages). A very recent study by Winter et al. (2013) compared two different statistical techniques for predicting the series to which each offense in the dataset belonged; Mokken scaling followed by discriminant function analysis and naïve Bayesian classifier analysis. However, in response to Bennell et al.'s criticism, ROC analysis was used to evaluate the accuracy of linking predictions. Winter et al.'s sample constituted 90 sex offenses committed by 30 serial sex offenders and 129 apparent one-off sex offenses from the UK. The AUCs ranged from .74-.89 with greater accuracy found when the one-off offenses were included in the dataset subject to analysis.

The remaining studies of the crime linkage principles with sex offenses have utilized ROC analysis to assess how accurately linked crime pairs can be distinguished from unlinked crime pairs. These studies have calculated a Jaccard's coefficient for each pair in their dataset which are then used to predict whether a given crime pair in the dataset is the work of the same offender (linked) or two different offenders (unlinked). If the two assumptions underpinning crime linkage are valid this discrimination task should be achieved with a high degree of accuracy (reflected by a large AUC). With a UK sample of 126 serious sexual assaults committed by 42 offenders, Bennell et al. (2009) found an AUC of .75 representing a good level of accuracy. Using the same sample but a different coding scheme with 36 variables (compared to the 27 variables used in the 2009 study), Bennell et al. (2010) reported an AUC of 0.81. More recently, with a South African sample of 22 serial sex offenders, Woodhams and Labuschagne (2012) reported AUCs ranging from .77-.88.

1.3. Limitations of Previous Linkage Research

While the studies outlined above provide some support for the assumptions underpinning crime linkage with samples of serial sex offenses, they have a number of limitations that have implications for the satisfaction of the rules of evidence outlined above. The first issue concerns the variation in how behavioral similarity is quantified and the statistical methods employed to test the principles of crime linkage. Disagreement in the approach to analysis presents difficulties when attempting comparisons between studies, or summation of the research. As noted above, Bennell and his colleagues (e.g., Bennell

et al., 2009; Bennell, et al., 2014) have made a convincing argument that ROC analysis should be adopted as the gold standard analysis since it is the only current method that is not impeded by threshold-specific results, producing a “pure measure of linking accuracy (i.e. the AUC)” (Bennell et al., 2009, p. 304).

ROC analysis was therefore adopted as the analytical technique in the current study.

Another limitation of existing crime linkage research with sex offenses is the sample sizes used. Most studies have sampled the crimes of 13 - 42 serial offenders ($N=42$ series (126 offenses), Bennell et al., 2009; $N = 16$ series (43 offenses), Santtila et al., 2005; $N = 30$ series (90 offenses), Winter et al., 2013; $N= 22$ series (119 offenses), Woodhams & Labuschagne, 2012). Yokota et al. (2007) is the exception to this with a sample of 188 serial offenders and 680 one-off offenders. The number of crimes included in statistical analyses in some previous studies has been further reduced by researchers adopting the practice of sampling a constant number of offenses from each series rather than utilizing all offenses in a series (e.g., Bennell et al., 2009, Santtila et al., 2005). As Woodhams and Labuschagne (2012) argued, police databases would not consist only of series of a constant number. The current study, therefore, sampled a larger number of serial sex offenders than most previous studies ($N = 50$ series) and included all known sex offenses ($N= 194$ offenses) comprising each series (thereby including series of differing length in the current study).

A further limitation is that most studies have used samples consisting solely of serial sex offenders (Bennell et al., 2009; Santtila et al., 2005; Woodhams & Labuschagne, 2012), whereas in reality crime linkage databases would contain both serial and one-off offenses. This is problematic for a number of reasons. First, it means previous research lacks realism and therefore the findings may not generalize to practice. Second, by including only serial offenders in an analysis, the researchers may be artificially separating the distributions of Jaccard’s coefficients for the linked and unlinked crime pairs (Woodhams, 2008). If it is accepted that serial sex offenders are consistent and distinctive in their crime scene behavior, in creating unlinked pairs from crime series, two crimes by two serial offenders with very different crime scene behavior are being paired. These should therefore look very dissimilar to one

another resulting in a small Jaccard's coefficient. This potential problem can be overcome by including one-off offenders in a sample. This also improves ecological validity.

Two recent studies have investigated if there are differences in the crime scene behavior of serial and one-off rapists. Corovic et al. (2012) stated that serial rapists were more criminally sophisticated than the one-off rapists in their first rapes. However, after applying a Bonferroni correction in the second part of their study, where they tried to predict if a rapist was a serial or one-off, the only predictors of a rapist's type were "kissed victim" (one-off offenders more likely to show this behavior) and "controlled victim" (more likely to be seen in serial offenders). Similarly, Slater et al. (2014) found that while there were some differences in the control and sexual behaviors between serial and one-off rapists, after a Benjamini-Hochberg correction, only one behavior remained significant, whether the victim was solicited by the offender (more likely by serial offenders). In summary, few differences have therefore been found between one-off and serial rapists meaning that it might be quite difficult to distinguish between linked crime pairs and unlinked crime pairs composed of one-off offenses.

To date, only a handful of studies have included one-off offenders in their samples (Grubin et al., 2001; Winter et al., 2013; Yokota et al., 2007), with only Winter et al. (2013) using ROC analysis. However, Winter et al.'s analysis utilized Mokken scaling, discriminant function analysis and naïve Bayesian classifier analysis to predict to which series an offense was likely to belong: this differs quite considerably from most research on the crime linkage assumptions that has adopted logistic regression and ROC analysis to determine how accurately the linkage status (linked versus unlinked) of crime pairs within a sample can be predicted. The current study adopted the more common method of assessing the accuracy of differentiating linked and unlinked crime pairs based on crime scene behavior using ROC analysis while still overcoming the other design flaws outlined. In addition to testing the validity of the underlying assumptions of crime linkage, this paper also considered Daubert criterion 3 (referring to the known error rate) by calculating figures for sensitivity and specificity.

Based on previous findings of studies of serial sex offending, it was hypothesized that linked crime pairs would be more similar in crime scene behavior than unlinked crime pairs, and that behavioral

similarity, as measured using Jaccard's coefficient, could accurately differentiate linked from unlinked crime pairs. No prediction was made as to the relative discrimination accuracy when using samples representing only serial sex offenses versus also including one-off sex offenses.

2. Method

2.1. Data

A national sample of stranger sex offenses was obtained from the Serious Crime Analysis Section (SCAS) of the National Crime Agency, UK. SCAS is a police analytical unit with national responsibility to carry out analytical work on behalf of all police forces. It collates and analyses information on serious crimes that fulfill its criteria, predominately stranger murders, and serious stranger sexual assaults and/or rapes. SCAS hold a database called the Violent Crime Linkage Analysis System (ViCLAS) which contains information about sexual crimes that meet their criteria including the behaviors displayed during each offense by the offender (National Crime Agency (SOCA), n.d.). They hold the most comprehensive dataset of stranger sex offenses in the UK.

A sample of sexual offenses committed by 50 serial and 50 one-off male offenders was provided by SCAS. Since the true ratio of serial to one-off offenders is unknown, it was decided that an even 1:1 ratio would best serve this study. Each crime was committed against a lone, adult, female victim by a lone stranger, adult, male offender. In total, the sample consisted of 194 serial sex offenses committed by the 50 serial offenders³ and 50 one-off sex offenses. The serial offenses spanned dates from 1977 to 2010, whilst the one-off offenses occurring between 1984 and 2009. The slight difference in dates is due to the request made to SCAS that sample collection start from the date of the request and extend backwards in time until 50 offenders of each type were identified. The definition of "serial" adopted in this study corresponds with international research programs on various forms of serial offending (e.g., Grubin et al., 2001; Santtila et al., 2005; Tonkin et al., 2008) and the Federal Bureau of Investigation's (2008) definition for serial murder, which is two or more crimes committed by the same individual. The term

³ Information as to how the serial offenses were originally linked (e.g. behavioral similarities, offender confessions, DNA) was not available to the researchers.

“one-off” is used within the paper to refer to apparent one-off offenses. It is impossible to know whether these are the crimes of sex offenders who genuinely committed only one offense, as established by a conviction, or whether they are in fact serial offenders whose other crimes have not been attributed to them. This point is returned to in the discussion of the paper. All the crimes included in the sample were from closed cases, where the offender had been convicted of the offense.

2.1.1. Serial offenders and their offenses. The 50 serial sexual offenders in the sample had a mean age at the time of offense of 31.69 years (Range = 18-58 years; $SD=8.76$). Seventy-two percent of the serial offenders were of White European ethnicity, the rest were of Dark European (8%), African or Caribbean (12%), Asian (2%), Arabic (2%), or Other ethnicity (4%). The mean series length was four offenses and the mode was three offenses (Range = 2-10 offenses). One hundred and two of the offenses were rapes, the rest were attempted rapes, assaults by penetration, indecent assaults, and indecent exposures (see Table 1). The remaining 31 offenses were other types of sexual offense (e.g., offenses that are classed as another type of crime but where a sexual component was evident). The inclusion of these unclassified sexual offenses for the serial offenders is a result of the sampling frame and the attempt to include all the sexual crimes in an offender’s series. The victims of the serial offenders had a mean age at the time of offense of 29.46 years (Range = 18-76 years; $SD=11.58$). Seventy-two percent of the victims were White European, with the rest being Dark European (1.5%), African or Caribbean (1%), Asian (1%), Arabic (1%), or Other (1.5%), and for 22% ethnicity was not recorded.

2.1.2. One-off offenders and their offenses. The sample of 50 one-off sexual offenders had a mean age at the time of the offense of 30.92 years (Range = 18-55 years; $SD=9.59$). Seventy percent of the offenders were of White European ethnicity, the rest were of Dark European (2%), African or Caribbean (12%), Asian (10%), Arabic (4%), or Other ethnicity (2%). From the 50 offenses they committed, 10 were classified as attempted rapes, while the other 40 were classified as completed rapes (see Table 1). The victims of the one-off offenders had a mean age at the time of offense of 28.20 years (Range = 18-59 years; $SD=9.51$). Eighty-four percent of the victims were of White European ethnicity,

with the rest being African or Caribbean (2%), or Other (4%), and for 10% their ethnicity was not recorded.

2.2. Procedure

The information about the offenders' crime scene behaviors for the offenses sampled was provided to the authors as a numerically coded spreadsheet which contained no personally identifying information. The spreadsheet contained 217 different modus operandi (MO) behaviors which included crime scene location descriptions, how the offender approached the victim, verbal themes, and sexual acts performed. Dichotomous coding was used for all the variables (1 = presence of a behavior during the offense, 0 = absence or unknown for a given behavior).

It was not possible for the authors to assess the inter-rater reliability of these data as the coding was completed before it was provided to the authors. However, all data inputted onto the SCAS database is completed in-house in a controlled environment by a team of highly trained individuals. Applicants are tested for their attention to detail and ability to identify relevant information prior to employment with SCAS. Recruits undergo several months of training before they are allowed to work autonomously, and only after they obtain a highly detailed knowledge of the system. To ensure accuracy and knowledge the training is rigorous and lengthy, and recruits must show clear understanding of behavioral vagaries. Initial training is not undertaken on a live database, and staff will not begin working on the live database until they have evidenced their capability to complete inputs accurately. Ensuring consistency in decision making in relation to difficult issues, a 'Quality Control Guide' is utilized by everyone inputting data on the database. Where an unusual aspect is encountered, for which no precedent has been set, a dedicated, experienced team meets to review the situation and make a decision. This decision is then recorded for future reference to ensure future consistency. Additionally, each inputted case goes through a detailed quality assurance process prior to any analysis taking place. This involves a review of the inputted information in comparison to case details, by an analyst from within the team and anomalies or errors are fed back to the inputter and amended on the database.

2.3. Analysis of data

As was outlined in the introduction to this article, Bennell and colleagues (Bennell & Jones, 2005; Bennell et al., 2009) have argued convincingly for the use of ROC analysis to test the assumptions of crime linkage. This study therefore followed a method pioneered by Bennell and Canter (2002), which has now been utilized in a substantial number of crime linkage studies (e.g., Bennell et al., 2009, 2010; Bennell & Jones, 2005; Tonkin et al., 2008; Woodhams & Labuschagne, 2012; Woodhams & Toye, 2007).

First, the Jaccard's coefficient was calculated for each crime pair in the dataset using a computer program created by Dr. Craig Bennell, called B-LINK^{4,5}. After calculating the Jaccard's coefficients the sample was split into linked and unlinked crime pairs. As outlined in the Introduction, traditionally studies have created linked and unlinked pairs from samples composed solely of serial offenses (hereafter referred to as the "traditional method"). This was also done in the current study to aid comparison with existing studies. In addition, sampling was extended to include one-off sex offenses allowing comparison of linked crime pairs with unlinked crime pairs containing offenses by one-off offenders (referred to hereafter as the "extended method"). These unlinked crime pairs were termed serial-serial, serial-one-off, and one-off-one-off pairs, respectively. The sample contained 365 linked crime pairs for both the traditional and extended methods. For the traditional method there were 18,356 unlinked crime pairs and in the extended method there were 29,281 unlinked crime pairs. Using either method, a significantly larger Jaccard's value for linked crime pairs compared to unlinked crime pairs would support the similarity and distinctiveness principles. This was assessed using Kruskal-Wallis tests because the distributions of the Jaccard's coefficients were significantly different to a normal distribution, as assessed by Kolmogorov-Smirnov tests.

⁴ The Jaccard's coefficients reported here were calculated based on all offense behaviors rather than categorizing behaviors into domains and then calculating a Jaccard's coefficient for each domain. This is because in existing studies (e.g., Bennell et al., 2009) as well as with this dataset, prediction using the coefficients generated from all offense behaviors was more accurate than that based on any individual domain. The output for individual domains can be obtained from the first author on request.

⁵ B-LINK (Bennell, 2002) calculates Jaccard's coefficient from dichotomously coded data about the crime scene behavior in each offense in the dataset. The output from B-LINK contains Jaccard's coefficient and information about whether each pair is linked or unlinked.

The Jaccard's data were also subject to a leave-one-out cross-validation (LOOCV) logistic regression analysis (Herrmann, 1998) which involves:

Taking each case out of the dataset one at a time. When a given case has been extracted, a logistic regression model is developed using the remaining dataset (representing the development data), which is then applied to the extracted case only (representing the validation data) to produce a predicted probability. This case is then returned to the dataset and the procedure repeated with the next case in the dataset until cases have been exhausted" (Woodhams & Labuschagne, 2012, p. 93).

This statistical procedure has been used in previous crime linkage studies (Tonkin, et al., 2012; Woodhams & Labuschagne, 2012). A Receiver Operating Characteristic (ROC) analysis was conducted on the predicted probabilities produced by the LOOCV logistic regression to assess how accurately using Jaccard's coefficients (behavioral similarity) crime pairs could be classified as linked or unlinked. Linkage status (linked or unlinked) was the state variable and the predicted probabilities were the test variable. These analyses were conducted with SPSS version 19.

Finally, Youden's index was calculated which represents the decision threshold (in this case, the Jaccard's coefficient) at which the proportion of hits is maximized while the proportion of false alarms is minimized (Bennell, 2005). This was calculated for both the traditional and extended sampling method. The equation for Youden's index is $J = pH + pCR - 1$, where pH is the probability of a hit and pCR is the probability of a correct rejection (Bennell & Jones, 2005). Having identified the Jaccard's coefficient at which the proportion of hits is maximized and the proportion of false alarms minimized, the sensitivity and specificity could be calculated, giving an indication of the error rate associated with adopting this statistical means of predicting linkage status. As applied to crime linkage, the sensitivity refers to the correct identification of linked crime pairs, and the specificity is the correct identification of unlinked crime pairs. Together sensitivity and specificity can be used to calculate the error rates, of Type I errors (false positives) and Type II errors (false negatives), which is the necessary information for Daubert criterion 3.

3. Results

3.1. Traditional Method of Testing Crime Linkage Principles (Series only sample)

The linked crime pairs had a median Jaccard's of .37 (Range = .04-1.00), whereas the unlinked crime pairs had a median of .16 (Range = .00-1.00). A Kruskal-Wallis test established that this difference was significant ($\chi^2(1, N=18721) = 595.50, p < .001$). The ROC analysis produced an Area Under the Curve (AUC) of .87 ($p < .001$, 95% CI = .85-.89) which represents an excellent level of predictive accuracy (Hosmer & Lemeshow, 2000). The ROC curve can be seen in Figure 1.

3.2. Extended Method of Testing Crime Linkage Principles (Series and one-off offenses sample)

The median of the linked crime pairs ($Mdn = .37$, Range = .04-1.00) indicated that they were more similar in crime scene behaviors than the unlinked crime pairs overall ($Mdn = .17$, Range = .00-1.00). This difference was significant ($\chi^2(1, N=29646) = 580.40, p < .001$). The AUC produced by the ROC analysis was .86 ($p < .001$, 95% CI = .84-.89) which represents an excellent level of discrimination accuracy. The ROC curve can be seen in Figure 2.

The median Jaccard's coefficient for each type of unlinked crime pair was also calculated: serial/serial unlinked ($Mdn=.16$), serial/one-off ($Mdn=.17$), and one-off/one-off ($Mdn=.20$). Unlinked pairs created by pairing the crimes of two different serial offenders are therefore the most dissimilar in crime scene behavior. A Friedman's test was conducted to compare the behavioral similarity of the three types of unlinked crime pair which produced a significant result ($\chi^2(2) = 94.30, p < 0.001$). Post hoc analysis using Wilcoxon signed-rank tests with a Bonferroni correction indicated there were significant differences in Jaccard's coefficients between all types of unlinked crime pairs,

When the Confidence Intervals (CI) for the AUCs produced using the traditional and extended method were compared they clearly overlap meaning that there is not a significant difference in discrimination accuracy (Bennell et al., 2009) whether distinguishing linked and unlinked pairs within a sample solely composed of sex offense series or within a sample comprising series and one-off sex offenses.

3.3. Decision-Making Thresholds

The Jaccard's coefficient that corresponded with the optimal threshold for the traditional sampling method was .245 and for the extended method was .241. The figures for specificity and sensitivity can be seen in Table 2. Under the traditional sampling condition, unlinked pairs were more accurately identified than linked pairs, as indicated by a larger proportion for specificity than sensitivity. The converse was true under the extended sampling condition although under both conditions the difference in size of proportions was only slight.

4. Discussion

In some jurisdictions, standards of admissibility (e.g., Daubert criteria and Rule 702) require the reliability of the assumptions of crime linkage to be empirically demonstrated and the error rate of crime linkage practice to be known. While several studies have attempted to assess these with serial sex offenses, methodological limitations affect their ecological validity and therefore the conclusions that can be drawn. The current study was designed to address these limitations by including one-off sex offenses in the dataset, ensuring a much larger sample of crime series than has been typical, and sampling all offenses from each known series.

In 2008, Woodhams cautioned against using only series-series unlinked pairs in analyses of the crime linkage assumptions for fear of inflating the difference in behavioral similarity between the linked and unlinked pairs. The findings from the current study give credence to her concerns: the series-series unlinked crime pairs were the most dissimilar in crime scene behavior. However, despite this, the linked crime pairs were significantly more similar in crime scene behavior than the unlinked crime pairs under both testing conditions (traditional and extended). This provides support for the underlying principles of crime linkage. However, it should be noted that the linked crime pairs were characterized by a wide range of Jaccard's coefficients. In theory, if crime scene behavior is driven by personal characteristics and less so by the situation, linked crime pairs should be characterized by high behavioral similarity. This certainly appeared to be the case for some of the pairings in the sample which had Jaccard's coefficients of 1.00 (indicating perfect similarity). On further investigation it became apparent that these perfectly similar linked crime pairs had relatively few "present" behaviors for comparison (i.e., six or fewer).

Commented [JW1]: Chelsea, you rightly note that you can calculate the false alarm and miss rate from the sensitivity and specificity. So, is it worth reporting the false alarm and miss rate for both conditions here for completeness?

Within the linked crime pairs there were other pairs where the similarity in behavior was virtually non-existent (with values of .04). Indeed, the distributions of Jaccard's coefficients for the linked and unlinked crime pairs overlapped quite considerably. These findings indicate that the principles of consistency and distinctiveness hold for some offense pairs better than others, which has implications both for the use of crime linkage in court but also for future research.

It is important that future research investigate the reasons for some offenders showing greater consistency than others. Possible avenues to explore include the psychological processes at work, specifically what is the function or meaning of the behavior to the offender. It would be expected that those offenders who enact certain behaviors that hold a strong function or purpose would be more likely to repeat that behavior. Alternatively, it may be that several different behaviors could all achieve the same, underlying function for the offender; hence, behavioral consistency would be less apparent, at least at the discrete behavioral level.

A more complete test of the principles underlying crime linkage was conducted using ROC analysis. Under conditions of traditional sampling, an AUC of .87 was found, representing an excellent level of discrimination accuracy. A similar figure (.86) was reported when sampling was extended to include apparent one-off sex offenses in the dataset. Both figures provide empirical support for the underlying principles of crime linkage; however, they also indicate that a degree of error still exists.

The AUCs found under both sampling conditions (.86-.87) are larger than the AUCs reported in existing studies of serial sex offenders (AUCs .75-.81) where researchers have artificially restricted the number of offenses sampled per series to a constant value (e.g., Bennell et al., 2009; Bennell et al., 2010; Woodhams & Labuschagne, 2012). However, they are similar to the figure of .88 reported by Woodhams and Labuschagne (2012) where they included all offenses from each series in the analysis ($N=119$ offenses), which was the sampling method adopted in the current study. As Woodhams and Labuschagne (2012) argued, it is more ecologically valid to sample all offenses from each series and it appears that artificially restricting the number of offenses per series sampled might underestimate the potential for

behaviorally linking crimes. Researchers are therefore urged to avoid restricting the number of crimes sampled per series in the future.

The inclusion of one-off sex offenses in the analysis had the opposite effect to what was found by Winter et al. (2013). Unlike Winter and colleagues, who found a slight improvement in discrimination accuracy when including one-off offenses, we found a slight decrease. This is likely attributed to the fact that in the current study the unlinked series-one-off pairs and the one-off-one-off pairs had larger Jaccard's coefficients than the series-series unlinked pairs. In this scenario it is more difficult to accurately distinguish linked from unlinked pairs because of the greater overlap in the Jaccard's distributions. However, the difference in AUC between the traditional and the extended method in this study was not found to be significant. Whether this would be the case if a larger sample of one-off offenses was used is unknown at present. It is possible that this effect would be accentuated with a larger sample of one-off offenses and this is something that future research needs to investigate.

The Youden's indexes were also very similar. Assuming that the findings from this sample would generalize to real world practice the police could apply the threshold to proactively screen for potentially linked crimes. The police could maximize hits while minimizing the number of false alarms if when crime pairs exceeded a Jaccard's coefficient of .241 they were classified as linked. In terms of the error rate in this simulation of crime linkage decision-making (Daubert criterion 3), adoption of this threshold would result in 79% of linked pairs (sensitivity) being correctly identified and 79% of unlinked pairs (specificity). The sensitivity and specificity for both the traditional and extended methods are similar to the findings of Winter et al. (2013). In their study, the sensitivity was 77.78 and the specificity was 82.62 for serial offenses only (traditional method) and 86.67 and 66.40 respectively for serial and one-off offenses (extended method). It is, however, important to emphasize that, despite the current study more closely simulating the conditions under which practitioners make crime linkage decisions, it does not fully reflect how such procedures would be expected to perform in real life where human beings rather than statistical models are predicting linkage status.

4.1. Limitations

The dataset used in this study more closely reflected what investigators would be working with during an investigation, and therefore made progress towards closing the gap between how crime linkage is researched and how it is conducted in the real world. However, it still has a number of limitations associated with the method which mean that it does not fully represent how the crime linkage principles might operate in practice and this is problematic when trying to produce research findings that would address criteria regarding admissibility.

As with much of the previous research in this area, the sample used here comprised the offenses of convicted offenders (Mokros & Alison, 2002; Santtila et al., 2005). The crime linkage research community has previously raised concerns about this method of sampling and how well findings generated from such samples would apply to reality where crime linkage is conducted on unsolved offenses (Bennell & Canter, 2002). However, a recent study by Woodhams and Labuschagne (2012) provides some reassurance since when comparing serial sex offense series first identified by the police (due to similar MO) to those first identified from DNA database hits, little evidence of a difference in behavioral similarity between the two types of series was found.

A further limitation, characteristic of any study that tries to sample one-off offenders, is that there can be no guarantee that those labeled as one-off offenders are not in fact serial offenders whose other offenses have yet to come to the attention of the police. Some crime linkage studies label crime series on the basis of arrest and this has been criticized with recommendations that criminal convictions are used as a basis instead (Snook, et al., in press). Despite the crime series in the current study being labeled on the basis of a conviction, it is still not possible to completely guarantee that the offenses included in a crime series are properly attributed because of the existence of miscarriages of justice. While none of these limitations can be overcome through improvements in design, it is still important that they be acknowledged.

It was not possible in the current study to know whether the analysts when coding the offenses onto the ViCLAS database were blind to series membership. It is therefore unknown whether the coding of crime scene behavior was influenced by knowledge of whether an offense belonged to a series or not.

However, a recent study by Pakkanen, et al. (2012), found that knowledge of the membership of an offense to a series did not appear to bias the coding of the behavioral data.

Finally, recent publications have queried the reliability of data contained within ViCLAS databases. Snook et al. (2012) tested the agreement between police officers coding a case file against the ViCLAS booklet. The study found only 10.38% of the variables reached acceptable levels of agreement, and the authors expressed concern about the implications for using such data for research. While these are valid concerns and indeed they warrant investigation, it is important to note that there are procedural differences between the coding of cases for ViCLAS in the UK compared to in Canada (where Snook et al.'s study was based). In Canada, the case files are coded against the ViCLAS booklet by police officers; whereas in the UK, a select number of trained assistant crime analysts do the coding. This is the main task associated with their role. There are also a number of quality assurance procedures in place as was outlined above. Therefore, while an explicit test of inter-rater reliability for ViCLAS coding in the UK is needed, findings from Snook et al.'s study are not directly applicable.

4.2. Conclusion

In line with the standards of admissibility operating in some jurisdictions (e.g., Daubert, Federal Rules of Evidence 702), much of the research in the crime linkage field has assessed the reliability of its underlying principles. The findings have been largely supportive; however, these studies have a number of limitations which impact ecological validity. The current study sought to address many of these by sampling a much larger set of stranger sex offenses than is typical, by sampling all offenses from each crime series, and by including the offenses of one-off sex offenders in the dataset. Linked and unlinked crime pairs could be discriminated with excellent levels of accuracy as assessed by ROC analyses therefore providing further support for the underlying principles of crime linkage (behavioral consistency and distinctiveness), but this time with a more ecologically valid sample. However, the results also show that predictions of linkage status are not without error.

References

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart.
- Bennell, C. (2002). *Behavioural consistency and discrimination in serial burglary* (Doctoral dissertation). Retrieved from OCLC WorldCat. (498749926).
- Bennell, C. (2005). Improving police decision making: General principles and practical applications of receiver operating characteristic analysis. *Applied Cognitive Psychology, 19*, 1157–1175. doi:10.1002/acp.1152
- Bennell, C., & Canter, D. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. *Science and Justice, 42*, 1-12. doi:10.1016/S1355-0306(02)71820-0
- Bennell, C., Gauthier, D., Gauthier, D., Melnyk, T., & Musolino, E. (2010). The impact of data degradation and sample size on the performance of two similarity coefficients used in behavioural linkage analysis. *Forensic Science International, 199*, 85-92. doi: 10.1016/j.forsciint.2010.03.017
- Bennell, C. & Jones, N. J. (2005). Between a ROC and a hard place: A method for linking serial burglaries by modus operandi. *Journal of Investigative Psychology and Offender Profiling, 2*, 23-41. doi:10.1002/jip.21
- Bennell, C., Jones, N. J., & Melnyk, T. (2009). Addressing problems with traditional crime linking methods using receiver operating characteristic analysis. *Legal and Criminological Psychology, 14*, 293-310. doi:10.1348/135532508X349336
- Bennell, C., Mugford, R., Ellingwood, H., & Woodhams, J. (2014). Linking crimes using behavioural cues: Current levels of linking accuracy and strategies for moving forward. *Journal of Investigative Psychology and Offender Profiling, 11*(1). 29-56. doi: 10.1002/jip.1395
- Bosco, D., Zappala, A., & Santtila, P. (2010). The admissibility of offender profiling in courtroom: A review of legal issues and court opinions. *International Journal of Law and Psychiatry, 33*, 184-191. doi:10.1016/j.ijlp.2010.03.009
- Canter, D. (1995). Psychology of offender profiling. In R. Bull, & D. Carson (Eds.), *Handbook of psychology in legal contexts* (pp. 343-355). Chichester, Hampshire : Wiley.

- Canter, D. (1996). A multivariate model of sexual offence behaviour: Developments in 'offender profiling'. In D. Canter (Ed.), *Psychology in action* (pp. 189-216). Hampshire, UK: Dartmouth Publishing Company.
- Charron, A. & Woodhams, J. (2010). A qualitative analysis of mock jurors' deliberations of linkage analysis evidence. *Journal of Investigative Psychology and Offender Profiling*, 7, 165-183. doi:10.1002/jip.119
- Corovic, J., Christianson, S. Å., & Bergman, L. R. (2012). From Crime scene actions in stranger rape to prediction of rapist type: Single-Victim or serial rapist?. *Behavioral Sciences and the Law*, 30(6), 764-781. doi: 10.1002/bsl.2026
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- Davies A (1992) Rapist's behavior: A three-aspect model as a basis for analysis and the identification of serial crime. doi:10.1016/03790738(92)90122-D
- Ellingwood, H., Mugford, R., Bennell, C., Melnyk, T., & Fritzon, K. (2013). Examining the role of similarity coefficients and the value of behavioural themes in attempts to link serial arson cases. *Journal of Investigative Psychology and Offender Profiling*, 10, 1-27. <http://dx.doi.org/10.1002/jip.1364>
- Federal Bureau of Investigation (2008). *Serial murder: multidisciplinary perspectives for investigators*. US Department of Justice, Washington.
- Federal Rules of Evidence (Rule 702), Pub. L. No. 93-595, §1, 88 Stat. 1937 (2011). <http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/2011%20Rules/Evidence%20Procedure.pdf>
- Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).
- Groscup, J.L., Penrod, S.D., Studebaker, C.A., Huss, M.T., & O'Neil, K.M., (2002). The effects of Daubert on the admissibility of expert testimony in state and federal criminal cases. *Psychology, Public Policy, and Law*, 8, 339-372. doi:10.1037/1076-8971.8.4.339

- Grubin, D., Kelly, P., & Brunson, C. (2001). Linking serious sexual assaults through behavior: HORS 215. *Home Office Research Development and Statistics Directorate*. Retrieved from <http://library.npia.police.uk/docs/hors/hors215.pdf>
- Hazelwood, R. R., & Warren, J. L. (2004). Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime. *Aggression and Violent Behavior, 9*, 307-318. doi:10.1016/j.avb.2004.02.002
- Herrmann, H. (1998). *Anwendung von Kreuzvalidierung und Bootstrap auf die logistische Regression und Diskriminanzanalyse*. Unpublished report.
- Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons.
- Kelly, L. (2010). The (in) credible words of women: False allegations in European rape research. *Violence against Women, 16*(12), 1345-1355. doi: 10.1177/1077801210387748
- Labuschagne, G. (2006). The use of a linkage analysis as evidence in the conviction of the Newcastle serial murderer, South Africa. *Journal of Investigative Psychology and Offender Profiling, 3*, 183-191. doi:10.1002/jip.51
- Labuschagne, G. (2012). The use of a linkage analysis as an investigative tool and evidential material in serial offenses. In K. Borgeson & K. Kuehnle (Eds.), *Serial offenders: Theory and practice* (pp. 187-215). Sudbury, Massachusetts: Jones & Bartlett Learning.
- Lin, S., & Brown, D. E. (2006). An outlier-based data association method for linking criminal incidents. *Decision Support Systems, 41*, 604-615. doi:10.1016/j.dss.2004.06.005
- Meyer, C. B. (2007). Criminal profiling as expert evidence? An international case law perspective. In R. N. Kocsis (Ed.), *Criminal profiling: International theory, research and practice* (pp. 207-247). Totowa, NJ: Humana Press. doi:10.1007/978-1-60327-146-2_10
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*(2), 246. doi: 10.1037/0033-295X.102.2.246

- Mokros, A. & Alison, L. J. (2002). Is offender profiling possible? Testing the predicted homology of crime scene actions and background characteristics in a sample of rapists. *Legal and Criminological Psychology*, 7, 25-43. doi:10.1348/135532502168360
- Pakkanen, T., Zappalà, A., Grönroos, C., & Santtila, P. (2012). The effects of coding bias on estimates of behavioural similarity in crime linking research of homicides. *Journal of Investigative Psychology and Offender Profiling*, 9(3), 223-234. doi: 10.1002/jip.1366
- Roberts, J. V., & Grossman, M. G. (1993). Sexual homicide in Canada: A descriptive analysis. *Annals of Sex Research*, 6, 5-25. <http://dx.doi.org/10.1007/BF00849743>
- Royal Canadian Mounted Police (n.d.). *Violent Crime Linkage System (ViCLAS)*. Retrieved January 25, 2013, from <http://www.rcmp-grc.gc.ca/tops-opst/bs-sc/viclas-salvac-eng.htm>
- Santtila, P., Junkkila, J., & Sandabba, N. K. (2005). Behavioural linking of stranger rapes. *Journal of Investigative Psychology and Offender Profiling*, 2, 87-103. doi:10.1002/jip.26
- Santtila, P., Pakkanen, T., Zappala, A., Bosco, D., Valkama, M., & Mokros, A. (2008). Behavioural crime linking in serial homicide. *Psychology, Crime and Law*, 14, 245-265. doi:10.1080/10683160701739679
- Serious Organised Crime Agency (SOCA). *Serious Crime Analysis Section (SCAS)*. Retrieved January 25, 2013, from <http://www.soca.gov.uk/about-soca/serious-crime-analysis-section/>
- Sjöstedt, G., Långström, N., Sturidsson, K., & Grann, M. (2004). Stability of modus operandi in sexual offending. *Criminal Justice and Behavior*, 31, 609-623. doi:10.1177/0093854804267094
- Slater, C., Woodhams, J., & Hamilton-Giachritsis, C. (2014). Can serial rapists be distinguished from one - off rapists?. *Behavioral sciences and the law*, 32(2), 220-239. doi: 10.1002/bsl.2096
- Snook, B., Luther, K., House, J.C., Bennell, C., & Taylor, P. (2012). The violent crime linkage analysis system: A test of interrater reliability. *Criminal Justice and Behavior*, 39, 607-619. doi: 10.1177/0093854811435208

- Snook, B., Luther, K., & MacDonald, S. (in press). Linking crimes with spatial behavior: A need to tackle some remaining methodological concerns. In J. Woodhams, & C. Bennell (Eds.), *Crime linkage: Theory, research and practice*. Boca Raton, FL: CRC Press.
- Tonkin, M., Grant, T., & Bond, J. W. (2008). To link or not to link: A test of the case linkage principles using serial car theft data. *Journal of Investigative Psychology and Offender Profiling, 5*, 59-77. doi:10.1002/jip.74
- Tonkin, M., Woodhams, J., Bull, R., & Bond, J. (2012). Behavioural case linkage with solved and unsolved crimes. *Forensic Science International, 222*, 146-153. doi:10.1016/j.forsciint.2012.05.017
- Winter, J. M., Lemeire, J., Meganck, S., Geboers, J., Rossi, G., & Mokros, A. (2013). Comparing the predictive accuracy of case linkage methods in serious sexual assaults. *Journal of Investigative Psychology and Offender Profiling, 10*, 28-56. doi:10.1002/jip.1372
- Woodhams, J. (2008). *Juvenile sex offending: An investigative perspective* (Doctoral dissertation).
- Woodhams, J., Bull, R., & Hollin, C. R. (2007). Case linkage: Identifying crimes committed by the same offender. In R.N. Kocsis (Ed.) *Criminal profiling: International theory, research and practice*. (pp. 117–133). Totowa, NJ: The Humana Press Inc.
- Woodhams, J. & Grant, T. (2006). Developing a categorization system for rapists' speech. *Psychology, Crime and Law, 12*, 245-260. doi:10.1080/10683160500151134
- Woodhams, J., Grant, T. D., & Price, A. R. G. (2007). From marine ecology to crime analysis: Improving the detection of serial sexual offenses using a taxonomic similarity measure. *Journal of Investigative Psychology and Offender Profiling, 4*, 17-27. doi: 10.1002/jip.55
- Woodhams, J., & Labuschagne, G. (2012). A test of case linkage principles with solved and unsolved serial rapes. *Journal of Police and Criminal Psychology, 27*, 85–98. doi: 10.1007/s11896-011-9091-1
- Woodhams, J., & Toye, K. (2007). An empirical test of the assumptions of case linkage and offender profiling with serial commercial robberies. *Psychology, Public Policy, and Law, 13*, 59-85.

doi:10.1037/1076-8971.13.1.59

Yokota, K., Fujita, G., Watanabe, K., Yoshimoto, K., & Wachi, T. (2007). Application of the behavioral investigative support system for profiling perpetrators of serial sexual assaults. *Behavioral Sciences and the Law*, 25, 841-856. doi:10.1002/bsl.793

Table 1

Types of Offenses Committed by Serial and One-Off Offenders

Offense	Serial Offenses N=194		One-Off Offenses N=50	
	<i>n</i>	%	<i>n</i>	%
Rape	102	52.6	40	80.0
Attempted Rape	18	9.3	10	20.0
Assault by Penetration	6	3.1	-	-
Indecent Assault	32	16.5	-	-
Indecent Exposure	5	2.6	-	-
Other Sexual Offense	31	16.0	-	-

Table 2

Overview of the indices associated with the optimal decision thresholds (Youden's index) for the traditional method of creating unlinked pairs (n = 18721) and the extended method (n = 29646).

	Sensitivity	Specificity
Traditional (series only)	78.90	80.97
Extended (series and one-off offenses)	79.45	78.63

Note: The equation for Youden's index is $J = p_H + p_{CR} - 1$, where p_H is the probability of a hit and p_{CR} is the probability of a correct rejection.

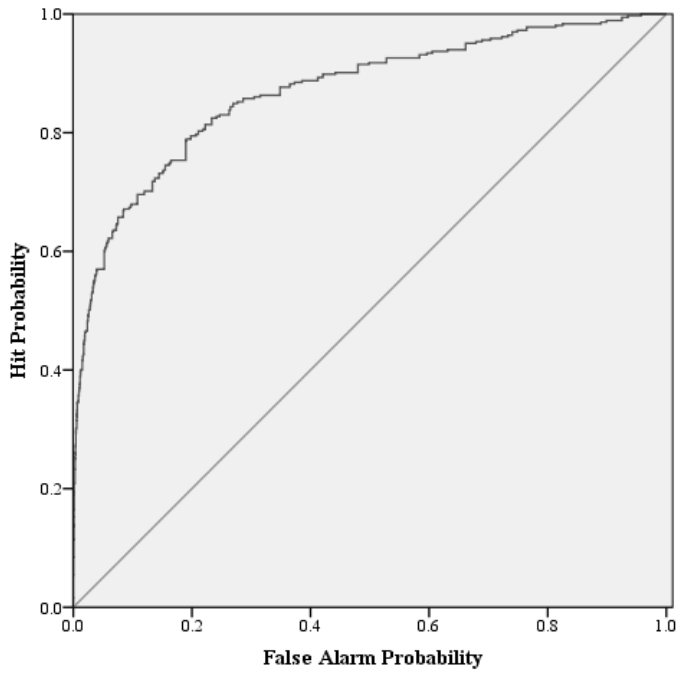


Figure 1. The ROC graph for the Traditional Method for differentiating linked and unlinked crime pairs using MO behaviors ($N = 18,721$). The Area Under the Curve (AUC), was .87 ($p < .001$, 95% CI = .85-.89).

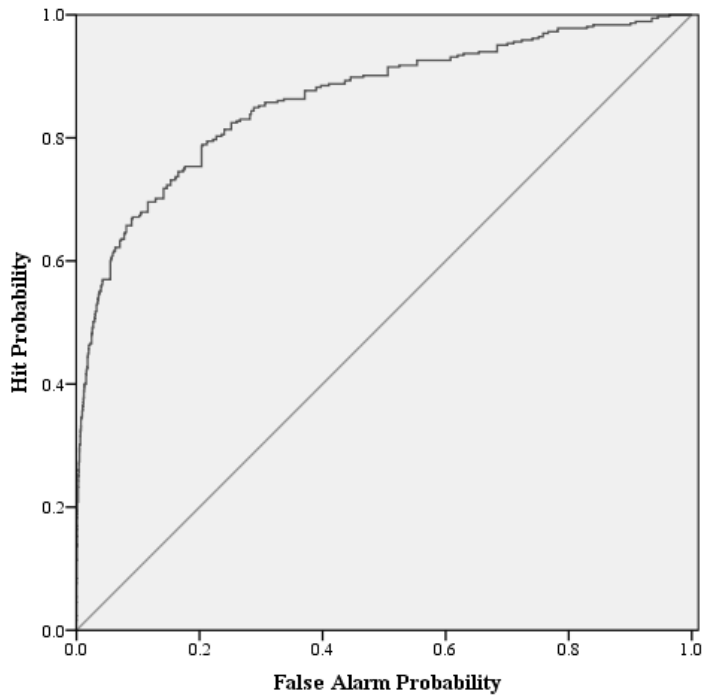


Figure 2. The ROC graph for the Extended Method for differentiating linked from unlinked crime pairs using MO behaviors ($N=29,646$). The Area Under the Curve (AUC) was .86 ($p<.001$, 95% CI= .84-.89).