Peer Reviewed
Received 5 January 2018
Received in revised form 9 March 2018
Accepted 27 March 2018
Available online 1 June 2018

Forensic Science Seminar
ISSN 2157-118X
Volume 8 Number 2
30 June 2018

Using Offender Crime Scene Behavior to Link Stranger Sexual Assaults: A Comparison of Three Statistical Approaches

Matthew Tonkin [a, *], Tom Pakkanen [b], Sirén, J. [c], Craig Bennell [d], Jessica Woodhams [e], Amy Burrell [f], Imre, H. [g], Jan Martin Winter [h, i], Lam, E. [h], ten Brinke, G. [h], Webb, M. [j], Labuschagne, G. N. [k], Leah Ashmore-Hills [l], Jasper Van der Kemp [m], Lipponen, S. [n], Rainbow, L. [j], Salfati, C. G. [o], Pekka Santtila [b]

a. Department of Criminology, University of Leicester, The Friars, 154 Upper New Walk, Leicester LE1 7QA, UK
b. Department of Psychology, Åbo Akademi University
c. Department of Biosciences, University of Helsinki, Finland
d. Department of Psychology, Carleton University
e. School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
f. Department of Psychology, Coventry University, Priory Street, Coventry CV1 5FB, UK
g. Belgian Federal Police, Brussels, Belgium
h. Dutch National Police, Zoetermeer, The Netherlands
i. Department of Clinical and Life Span Psychology (KLEP), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
j. National Crime Agency, UK
k. L&S Threat Management, South Africa
l. Department of Psychology, Birmingham City University, 4 Cardigan Street, Birmingham B4 7BD, UK
m. VU University Amsterdam, Faculty of Law, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
n. Finnish National Police, Helsinki, Finland
o. Department of Psychology, John Jay College of Criminal Justice, City University of New York, New York, NY 10019, USA
*CORRESPONDING AUTHOR. E-mail: mjt46@le.ac.uk

ABSTRACT *Purpose*: This study compared the utility of different statistical methods in differentiating sexual crimes committed by the same person from sexual crimes committed by different persons. *Methods*: Logistic regression, iterative classification tree (ICT), and Bayesian analysis were applied to a dataset of 3,364 solved, unsolved, serial, and apparent one-off sexual assaults committed in five countries. Receiver Operating Characteristic analysis was used to compare the statistical approaches. *Results*: All approaches achieved statistically significant levels of discrimination accuracy. Two out of three Bayesian methods achieved a statistically higher level of accuracy (Areas Under the Curve [AUC] = 0.89 [Bayesian coding method 1]; AUC = 0.91 [Bayesian coding method 3]) than ICT analysis (AUC = 0.88), logistic regression (AUC = 0.87), and Bayesian coding method 2 (AUC = 0.86). *Conclusions*: The ability to capture/utilize between-offender differences in behavioral consistency appear to be of benefit when linking sexual offenses. Statistical approaches that utilize individual offender behaviors when generating crime linkage predictions may be preferable to approaches that rely on a single summary score of behavioral similarity. Crime linkage decision-support tools should incorporate a range of statistical methods and future research must compare these methods in terms of accuracy, usability, and suitability for practice.

KEY WORDS Crime linkage, Comparative case analysis, Bayesian analysis, Logistic regression, Classification tree analysis, Stranger sexual assault, Forensic science

I. INTRODUCTION

One of the most well documented findings in criminology is that the majority of crime is committed by a minority of serial offenders who impose significant costs on society (e.g., Piquero, Farrington, & Blumstein, 2007). Estimates suggest, for example, that 6 - 10% of offenders are responsible for more than half of all crime committed in the United States (US) and the United Kingdom (UK) (Dodd, Nicholas, Povey, & Walker, 2004; Wolfgang, Figlio, & Sellin, 1972), with the average career criminal costing society more than $1.14 million during their lifetime (DeLisi & Gatling, 2003). Developing methods for catching and convicting serial offenders is, therefore, a significant priority for the criminal justice system.

To tackle serial offending effectively, methods must be developed to identify so-called linked crime series, which consist

of two or more crimes that have been committed by the same offender or the same group of offenders (Woodhams, Hollin, & Bull, 2007). In the absence of physical trace material (e.g., DNA) to link crime scenes, it has been suggested that similarity in offender crime scene behavior might be used (e.g., Bennell & Jones, 2005; Burrell, Bond, & Bull, 2012; Woodhams & Toye, 2007). The assumption is that crimes committed in a similar way behaviorally (e.g., using a similar level and type of violence, similar methods of controlling the victim, etc.) might be categorized as linked (i.e., committed by the same person) whereas crimes constituting very different behavior might be categorized as unlinked (i.e., committed by different persons) (Bennell & Canter, 2002). This procedure has been referred to using various names, including crime linkage, behavioral case linkage, comparative case analysis, and crime linkage analysis [1]. The term crime linkage will be used throughout the current article.

If crimes can be accurately linked, crime linkage affords a number of potential benefits to criminal justice agencies. First, it allows the evidence collected across several investigations to be pooled, which can increase the quantity and quality of evidence available with which to catch and convict serial offenders (Grubin, Kelly, & Brunsdon, 2001). Second, the ability to link multiple crimes to a single offender enables the police to combine different investigations, thereby helping to avoid duplication of roles, responsibilities, and investigative work that would occur if these crimes were investigated separately (Woodhams, Hollin et al., 2007). Ultimately, this creates a more efficient and streamlined investigative process (Woodhams, Hollin et al., 2007), which is of significant benefit at a time when law enforcement agencies are facing considerable budget cuts and resource constraints. Third, when crimes are successfully linked, it has been suggested that each individual victim gains confidence and credibility from the others, thereby increasing the likelihood that cases will successfully reach court (Davies, 1992). This is particularly important for sexual crimes where it is estimated that only six out of every 1000 rapists in the US will be incarcerated and high levels of attrition are reported at all levels of the criminal justice process [2].

Given these potential benefits, it is unsurprising that law enforcement units have been established around the world to facilitate the behavioral analysis of crime (including crime linkage). For example, such units have been established in the UK, the US, Canada, Belgium, the Netherlands, South Africa, Germany, Japan, New Zealand, France, the Czech Republic, and Switzerland (to name but a few countries). For a variety of reasons, however, the task of crime linkage is a considerable challenge for criminal justice practitioners. Crime linkage involves a number of analytical steps (as outlined by Woodhams, Bull, & Hollin, 2007), including identifying the offender behaviors present in a given crime (of which there can be many; Bennell, Bloomfield, Snook, Taylor, & Barnes, 2010), identifying behavioral similarities and differences across multiple crimes, considering situational circumstances and base rates [3], and then summarizing this information in a written and/or verbal report. This process might involve sifting through hundreds, possibly thousands, of crimes to identify offenses that share similar offender behaviors [4]. Crime linkage is, therefore, a process that can be very time-consuming and can place considerable cognitive load on criminal justice practitioners (Santtila, Korpela, & Häkkänen, 2004).

One approach to overcoming (or at least partially addressing) the challenges associated with crime linkage is to develop computerized decision-support tools that can analyze vast quantities of crime scene information in a quick and efficient manner. These tools would then provide the practitioner with a prioritized list of potentially linked crimes for further investigation/analysis and a simple, easy-to-process summary of the behavioral similarities and differences between these various crimes (e.g., Canter & Youngs, 2008; Grubin et al., 2001; Oatley, Ewart, & Zeleznikow, 2006; Woodhams, Bull et al., 2007). Such tools might help to reduce the cognitive load on analysts when they are conducting crime linkage, which would be beneficial because excessive load has been shown to hamper performance and lead to decision-making errors in a variety of contexts (e.g., see Adcock, 2000, for a review). Furthermore, computerized decision-support tools that automate certain parts of the analytical process might increase the efficiency of crime linkage units, allowing them to analyze more cases (in less time) than they are currently able to. This would help criminal justice agencies to continue meeting operational demand despite decreasing resources.

Over the last decade, a growing body of research has sought to develop statistical methods that might underpin computerized

1 It is important to note that often these terms are used inter-changeably, but some scholars use these terms to refer to distinctly different analytical processes (see Rainbow, 2015).

2 This estimate is based on a range of sources summarized at: https://www.rainn.org/statistics/criminal-justice-system

3 In judging whether a behavioral similarity/difference is useful for determining crime linkage status (linked/unlinked), the practitioner must consider situational circumstances. For example, apparent behavioral differences between two crimes might be explained by the fact an offender was interrupted in one crime but not the other, and the interruption forced the offender to alter his/her behavior. In which case, the differences might not be considered that useful by the practitioner. When considering whether a given behavioral similarity is useful, the practitioner must consider base rate information indicating how frequently given behaviors occur in a particular type of crime. That is, it is perhaps not that useful if the behaviors shared across two crimes consist only of behaviors that are very common to that particular type of offense (e.g., vaginal penetration from the front is common in sexual offenses; Santtila, Junkkila, & Sandnabba, 2005; Winter et al., 2013). It is much more useful if relatively rare behaviors are shared across several crimes, and in such a situation it would be more likely that one would conclude that the two crimes were linked.

4 For example, the unit responsible for conducting crime linkage with sexual offenses in the UK, the Serious Crime Analysis Section (SCAS), hold a database containing over 25,000 offenses within which their analysts must search for potentially linked crimes.

crime linkage support tools (e.g., Bennell & Jones, 2005; Burrell et al., 2012; Ellingwood, Mugford, Bennell, Melnyk, & Fritzon, 2013; Santtila, Junkkila, & Sandnabba, 2005; Santtila et al., 2008; Tonkin, Grant, & Bond, 2008; Winter et al., 2013; Woodhams & Labuschagne, 2012; Woodhams & Toye, 2007; Yokota, Fujita, Watanabe, Yoshimoto, & Wachi, 2007). These studies have found support for the two theoretical assumptions that underpin crime linkage (behavioral consistency and distinctiveness [5]) and have demonstrated moderate to high levels of accuracy when using offender crime scene behavior to distinguish between linked and unlinked offenses (see Bennell, Mugford, Ellingwood, & Woodhams, 2014, for a review). Within this literature, a range of statistical methods have been explored, including (but not limited to) logistic regression, classification tree analysis, and Bayesian analysis. There are, however, very few studies that have drawn direct comparisons between different statistical approaches. Consequently, it is not possible to determine from existing literature which (out of the many available statistical methods; Bennell, Goodwill, & Chinneck, 2015) is the most suitable/offers the greatest potential for supporting the development of computerized crime linkage decision-support tools. Ultimately, this is preventing researchers from developing evidence-based tools, thereby limiting the value of existing research to criminal justice practitioners.

The current study aims to overcome this fundamental limitation by comparing a variety of statistical methods in terms of their ability to distinguish between linked and unlinked crimes (referred to hereafter as discrimination accuracy). This follows a methodology originally developed by Bennell (2002), which has since been adopted in numerous peer-reviewed studies (e.g., Bennell & Jones, 2005; Burrell et al., 2012; Ellingwood et al., 2013; Tonkin et al., 2008; Woodhams & Toye, 2007). The methodology involves creating linked crime pairs (which contain two crimes committed by the same offender) and unlinked crime pairs (which contain two crimes committed by different offenders). A statistical measure is then calculated indicating the behavioral similarity between the two crimes in each pair (based on a range of offense behaviors, such as whether a victim was tied up, what type of violence was perpetrated, and so on). These similarity coefficients are then entered into different statistical analyses (e.g., logistic regression, classification tree analysis) and used to generate predictions as to whether the crime pairs are linked or not. The accuracy of these predictions is then evaluated (typically using Receiver Operating Characteristic (ROC) analysis, which is described in the Method section of this paper). In Bennell's methodology, the creation of unlinked crime pairs tests whether there are differences between offenders when offending (behavioral distinctiveness) and the creation of linked crime pairs tests whether offenders repeat elements of their offending behavior from one crime to the next (behavioral consistency). Thus, if the crime linkage principles of behavioral consistency and distinctiveness are shown to have support, we would expect linked crime pairs to be more behaviorally similar than unlinked pairs. By comparing different statistical methods in terms of discrimination accuracy, this indicates which method is best able to capture behavioral consistency and distinctiveness and to subsequently use that information to predict whether crimes are linked or not.

Not only does this methodology test the underlying theoretical assumptions of crime linkage, but it also relates to the various crime linkage tasks facing criminal justice practitioners (see Rainbow, 2015; Woodhams, Bull et al., 2007). Within the literature, three different crime linkage scenarios are commonly described. In scenario 1 the practitioner is presented with an index crime and asked to find other offenses within a large database that might be linked to that particular index offense. This scenario has been referred to as comparative case analysis (Rainbow, 2015) and reactive case linkage (Woodhams, Bull et al., 2007). In scenario 2 the practitioner searches through a large database to find linked offenses without comparison to a specific index crime. This task has been referred to in the literature as proactive case linkage (Woodhams, Bull et al., 2007). In the third scenario the practitioner is presented with a predefined set of crimes (10 crimes in this example) and is asked to decide whether the crimes are linked or not. This task has been referred to as crime/case linkage analysis in the literature (Rainbow, 2015). While these three scenarios differ, each task can be broken down into a series of pairwise comparisons. In the first scenario, the most comprehensive way to address such a task would be to compare the index crime to every single crime in the database, with the most behaviorally similar offenses highlighted as the most likely to be linked. This would involve creating a large number of pairwise comparisons (e.g., between the index crime and crime 1 in the database, between the index crime and crime 2 in the database, and so on). In the second scenario, the most comprehensive way of addressing this task would be to remove each crime in the database one at a time and then compare that crime to those crimes remaining in the database. When all pairwise comparisons have been made, this crime would be returned to the database and the next crime removed and compared in the same pairwise fashion to all remaining crimes. Again, the most behaviorally similar offenses would be highlighted as those most likely to be linked. In the third scenario, the same approach as that taken for scenario 2 could be utilized. For example, if there were 10 crimes in the predefined set, crime 1 would be removed and compared individually to crimes 2, 3, 4, and so on. When these pairwise comparisons were completed, crime 2 would be

5 In order for crime linkage to function reliably and accurately, offenders must repeat certain elements of their offending behavior from one offense to the next (behavioral consistency) and there must be individual differences between offenders in the way that they commit crime (behavioral distinctiveness), otherwise it will not be possible to distinguish the crimes of one offender from those of another (Woodhams, Hollin et al., 2007).

compared to crimes 3, 4, 5, and so on. This process would be repeated until all pairwise comparisons had been made. The practitioner could then plot these crimes (based on the similarity scores produced by the pairwise comparisons) and if the crimes clustered together this would suggest that they were linked [6]. Thus, all three crime linkage tasks described above can be addressed by creating multiple pairwise comparisons. Bennell's methodology directly replicates this process and tests which statistical approaches are best able to distinguish between linked and unlinked crime pairs. This provides an insight into which statistical methods have the greatest potential for supporting the development of computerized crime linkage decision-support tools.

The current study will compare binary logistic regression, iterative classification tree (ICT) analysis [7], and Bayesian analysis. As noted above, while a number of studies have examined these approaches individually, very few studies have compared them in terms of discrimination accuracy. In fact, there is just one published study to the authors' knowledge that has compared all three statistical approaches using the same dataset (Porter, 2014). In that study, Porter found a comparable level of discrimination accuracy when using boosted trees (a form of classification tree analysis), Naïve Bayes, and logistic regression models to distinguish between linked and unlinked breaking and entering offenses from Baltimore County, US. Likewise, there were few statistically significant differences in discrimination accuracy when comparing logistic regression and ICT analysis (Bayesian analysis was not investigated) using samples of residential burglary from Finland, car thefts from the UK, and adult stranger rapes from Canada (XXXX, under review-anonymized for peer review; Tonkin, Woodhams, Bull, Bond, & Santtila, 2012).

The fact that similar levels of discrimination accuracy were observed across the statistical methods tested in these studies is somewhat surprising because regression, Bayesian, and classification tree analysis adopt very different approaches to generating crime linkage predictions. Users of binary logistic regression have tended to adopt (what has been referred to in the literature as) a 'one-size-fits-all' approach, meaning that a single statistical algorithm is developed for predicting linkage status (linked/unlinked) and this approach is then applied to all cases (Tonkin et al., 2012). Consequently, the exact same offender behaviors are used in the same way to generate predictions across all crimes in a given dataset. Such an approach has been criticized because it is not consistent with findings suggesting that behavioral consistency is differentially expressed from one offender to the next (e.g., some offenders might be consistent in sexual behaviors, whereas other offenders might be consistent in control behaviors) (Grubin et al., 2001). The one-size-fits-all approach adopted by logistic regression would not capture such nuances in offender behavior.

In response to these criticisms, classification tree analysis has been proposed as an alternative (and arguably more appropriate) statistical approach for generating crime linkage predictions (XXXX, under review-anonymized for peer review; Tonkin et al., 2012). One reason for the presumed superiority of classification tree analysis is because it allows for different predictive methods (i.e., different combinations of offender behavior) to be used for different sub-groups of offenders/crimes (Steadman et al., 2000), thereby allowing some idiographic flexibility in decision-making that is not possible with logistic regression.

There are, however, limitations to the way in which both logistic regression and classification tree analysis have been utilized in the literature. Typically, a single coefficient is calculated to indicate how behaviorally similar the two crimes are in each linked and unlinked pair (which involves combining the information contained across multiple behavioral variables, e.g., the offender wore a mask- yes/no; the victim was gagged- yes/no; and so on). This number is subsequently used in the regression/classification tree analysis to predict whether crime pairs are linked or not. This approach is problematic, though, because information is lost by combining multiple behavioral variables into a single similarity value. For example, crime pair A and crime pair B might both receive a similarity score of 0.25, which means they are treated for the purposes of logistic regression and classification tree analysis as the same in terms of their behavioral similarity. But, the value of 0.25 tells us nothing about which particular behaviors were similar (and not similar) across the two crimes. Indeed, while crime pairs A and B might have the same similarity score, the specific shared behaviors that contributed to producing this score could be completely different. Such a loss of information can, however, be avoided using other statistical approaches, such as Bayesian analysis, which use the individual behavioral variables to generate crime linkage predictions (rather than relying on a single summary score of behavioral similarity). Consequently, Bayesian-based analyses are a potentially very useful family of techniques for exploring whether and how crimes can be linked using offender crime scene behavior.

Given the above, the current study compares logistic regression, ICT analysis, and Bayesian analysis in terms of their ability to use offender crime scene behavior to distinguish between linked and unlinked sexual crimes [8]. While crime linkage is conducted in

6 Similar to smallest space analysis and other multidimensional scaling procedures (e.g., Santtila et al., 2005).

7 Please refer to the Method section of this paper for a description of iterative classification tree analysis, which is distinct from classification tree analysis where only one tree is constructed and used to make classification decisions (compared to constructing multiple trees).

8 It should be noted that the statistical methods tested in the current study differ from those tested by Porter (2014). For example, Porter (2014) uses boosted trees whereas the current study relies on the Chi-squared Automatic Interaction Detector (CHAID) algorithm available in PASW (see the Analytic Strategy section). Please contact the authors if you wish to further discuss similarities

practice with a range of crime types (including both person-oriented offences, such as rape and homicide, and property-oriented offences such as burglary, robbery, and car theft), a focus on sexual offences is justified because these crimes have particularly significant emotional, psychological, and health consequences for victims (e.g., see Rentoul & Appleboom, 1997; Resick, 1993). Furthermore, sexual offences are estimated to have the second largest financial cost for society (behind homicide), considering a range of costs for victims, the criminal justice system, and wider society (McCollister, French, & Fang, 2010). As far as the authors are aware, none of the statistical methods tested in this study are currently used by criminal justice practitioners to link crimes in practice, but (based on previous research) they all appear to offer some potential in this regard. It is hypothesized that discrimination accuracy will be greatest using the Bayesian-based statistical approaches due to the loss of information that can occur when using logistic regression and ICT analysis (as discussed previously). To facilitate these comparisons, a sample comprising over 3,000 sexual offenses committed in five countries is collated, which represents the largest, most diverse, and most ecologically valid dataset ever collected to investigate crime linkage with sexual offenses. This study, therefore, provides a unique insight into which statistical approach offers the greatest potential for supporting the development of crime linkage decision-support tools. This fills an important gap in the crime linkage literature (because such comparisons have never before been made) and provides a key step towards translating these findings into a usable tool that can enhance law enforcement practice.

II. METHOD

Data

The study utilized police crime data relating to 3,364 stranger sexual offenses committed by 3,018 offenders (mean number of sexual offenses per series = 3.25, range = 2 – 32 crimes). These data were provided by law enforcement agencies from five countries: 1) the Serious Crime Analysis Section (SCAS, UK, n = 2,579 offenses); 2) the South African Police Service (SAPS; n = 245 offenses); 3) the Finnish National Police (n = 123 offenses); 4) the Dutch National Police (n = 173 offenses); and 5) the Belgian Federal Police (n = 244 offenses). Within these data, there were solved serial crimes (n = 2,081), unsolved serial crimes (n = 92), and solved apparent one-off crimes (n = 1,191). In this study, unsolved crime series consisted of crimes that had been linked via DNA. Thus, while they remain unsolved, we can be somewhat confident that the same offender was responsible (this is important because otherwise we will not know whether the predictions generated by our statistical methods are accurate or not). Apparent one-off crimes consisted of crimes committed by an offender who only had one recorded conviction for sexual offending at the time of data collection. The inclusion of unsolved and apparent one-off crimes was important because, when practitioners are searching for linked crimes in practice, the databases they search contain a mixture of solved, unsolved, serial, and one-off offenses. By including such offenses in our research, this helped to ensure that the findings were more ecologically valid than those produced in the majority of previous research (which failed to include unsolved and one-off offenses; e.g., Bennell & Jones, 2005; Burrell et al., 2012; Ellingwood et al., 2013; Santtila et al., 2005, 2008; Tonkin et al., 2008; Woodhams & Toye, 2007).

For each crime in the dataset, information pertaining to 166 binary behavioral variables was collated. This encompassed a range of offender behaviors, including: 1) control behaviors, consisting of behaviors designed to gain control over the victim and offending situation (e.g., weapon use, use of violence, etc.); 2) escape behaviors, designed to help the offender evade capture or exit the crime scene (e.g., wearing gloves or a disguise, taking forensic precautions); 3) style behaviors, which are not directly necessary for the offense to be successfully completed (e.g., the offender complimenting the victim); 4) sexual behaviors (e.g., whether the victim was penetrated and how, etc.); and 5) target selection variables (e.g., the time and day of the offense, the age and gender of the victim, etc.).

These data were collated from the five countries in a number of ways. The Finnish data were collated from two pre-existing research datasets (Häkkänen, Lindlöf, & Santtila, 2004; Santtila et al., 2005). Inter-rater reliability (IRR) for the original datasets are published in the respective papers (a mean Cohen's Kappa of 0.77 for Santtila et al. (2005) and only variables with $K > 0.61$ (with two exceptions) were kept by Häkkänen et al. (2004)). The South African data were collected by the sixth author who coded behaviors displayed in rape cases from the hard copy case files provided by the SAPS. A coding dictionary was developed in collaboration with our practitioner partners in other countries to ensure comparable data would be collected. The first five series (n = 20 cases) were dual coded by the fifth and sixth authors and IRR analysis performed. Where low scores were achieved ($K < 0.60$) a decision was made to either remove these from the dataset (n = 10 variables) or retain with a clarification of the coding definition (n = 9 variables). The IRR process also resulted in the clarification of variable definitions for a further 15 variables. Finally, the IRR resulted in collapsing two variables into one (minimal and moderate violence) due to low IRR scores ($K = 0.56$ and -0.07 respectively).

The remaining three datasets (UK, Belgium, and the Netherlands) were collated from data stored on the Violent Crime Linkage Analysis System (ViCLAS; see Collins, Johnson, Choy, Davidson, & MacKay, 1998). ViCLAS is a database that stores records of serious crimes (typically stranger sexual offenses and sexual homicides) including the crime scene behavior engaged in

and differences between those methods used in the current study and those used by Porter (2014).

by the offender. It has the functionality to be interrogated for crimes which share behavioral characteristics and is used to support the process of crime linkage in Belgium, the Czech Republic, France, Germany, Ireland, the Netherlands, New Zealand, Switzerland and the United Kingdom (Wilson & Bruer, 2017). In the UK, Belgium, and the Netherlands, police investigators submit case papers to the analytical units and the data are entered on to ViCLAS by trained analysts within these units. The training of analysts is a lengthy process typically lasting several months (but it can last as long as a year, or longer if necessary) and involving close supervision by an experienced senior analyst. Data entry on to ViCLAS is closely supervised by senior analysts and guided by a detailed quality control guide/coding manual, which explains the meaning of individual ViCLAS variables and gives examples of how these variables should and should not be coded. Consequently, all analysts entering data on to the ViCLAS system are following the same coding rules. Furthermore, before analysis begins on any case, that case is reviewed to ensure that the information entered onto the ViCLAS system matches the original police files. Any inconsistencies are fed back to the analyst who entered the data on to the system and amended within ViCLAS itself.

For the purposes of gathering data for the current study, an analyst from SCAS extracted the UK data directly from ViCLAS. In Belgium and the Netherlands, crime analysts manually extracted data from ViCLAS and other relevant systems (e.g., crime records to identify solved and unsolved cases). In the Netherlands, all data retrieved from ViCLAS was reviewed by the analysts against the original paper files to ensure the coding was in accordance with the current coding dictionary and quality control was assessed using the current manual. These datasets were anonymized, encrypted and sent to the research team.

Once all five datasets had been received, these were reformatted into one row per offense and manually joined together by the sixth and thirteenth authors. The individual datasets contained a range of behavioral variables, and those which overlapped were retained for the project. Variable matching was completed manually using variable labels with input from the practitioner partners to ensure matched variables represented behaviors that were as similar as possible across all five countries. The liaison with practitioner partners was essential, as data agreements did not permit academic partners to have sight of coding dictionaries for ViCLAS countries.

Analytic Strategy

Following a method developed by Professor Craig Bennell and used by many crime linkage researchers since (see Bennell et al., 2014, for a review), the first stage of the analysis involved generating linked and unlinked crime pairs from the data. Linked crime pairs contained two crimes committed by the same offender and unlinked crime pairs contained two crimes committed by different offenders. All possible linked and unlinked crime pairs were created from the data, resulting in a sample of n = 4,569 linked crime pairs and n = 5,651,997 unlinked pairs. Once these pairs had been created, different analytical procedures were used for the binary logistic regression, ICT analysis, and Bayesian analysis.

Binary logistic regression analysis.

First, a Jaccard's coefficient was calculated for each linked and unlinked crime pair to provide a measure of how similar the two crimes were in terms of offender crime scene behavior (based on the 166 binary behavioral variables mentioned previously). Jaccard's coefficient was calculated using the following formula: J = a ÷ (a + b + c), where *J* refers to the Jaccard's coefficient, *a* the number of behaviors present in both crimes in the pair, *b* the number of behaviors present in crime one but absent from crime two, and *c* the number of behaviors absent in crime one but present in crime two. Jaccard's coefficient is one of many similarity coefficients that can be used with binary data, and has been utilized in numerous crime linkage studies (e.g., Bennell & Canter, 2002; Burrell et al., 2012; Tonkin et al., 2008; Woodhams & Toye, 2007). The coefficient can range from 0 (indicating that none of the behaviors analyzed were present in both crimes in the pair) to 1.00 (indicating that the exact same behaviors were present in both crimes).

The Jaccard's coefficient was entered as an independent variable in the logistic regression analysis, with the aim of building a statistical model that could predict the likelihood of a crime pair being linked. Given that the ultimate aim of crime linkage research is to generate methods that might be used to link future cases (beyond the sample studied), it is vital that methods of cross-validation are used. The leave-one-out classification method (LOOCV) was used in the current study to cross-validate the logistic regression model. The LOOCV method involved removing each crime pair from the sample one at a time, and the remaining data were then used to develop a logistic regression model. This regression model was subsequently applied to the extracted pair to produce a predicted probability value (ranging from 0, indicating a low predicted probability of the crime pair being linked, to 1.00, indicating a high predicted probability of the pair being linked). This pair was then returned to the dataset and the procedure repeated with the next pair until a probability value had been calculated for all linked and unlinked crime pairs in the sample (Woodhams & Labuschagne, 2012). These predicted probability values were used in subsequent analysis to test the discrimination accuracy of the regression model (as described in more detail below).

Iterative classification tree (ICT) analysis.

For the ICT analysis, the 166 behavioral variables were split into five types of offender behavior (as described in the Data section of this paper). These so-called behavioral domains were based on those utilized in previous research (e.g., Bennell, Gauthier, Gauthier, Melnyk, & Musolino, 2010; Grubin et al.,

2001; Woodhams, Grant, & Price, 2007). As discussed in the introduction, one of the proposed advantages of classification tree analysis over logistic regression is that it can more easily capture/utilize differences between offenders in how they display behavioral consistency when offending. This will only be achieved, however, if the analysis breaks offender behavior down into different types. If there is just a single, combined measure of offender behavior then only one strategy for linking offenses would emerge from the analysis (i.e., the ICT analysis would produce a one-size-fits-all model). However, by identifying separate behavioral domains, this allows different linking strategies to be developed for different sub-groups of sexual offenders (e.g., linkage decisions might be generated using control and escape behaviors for one sub-group of offenders but for a different sub-group target selection and sexual behaviors might be used). The five behavioral domains utilized in the current study were: 1) control behaviors; 2) escape behaviors; 3) style behaviors; 4) sexual behaviors; and 5) target selection variables. Jaccard's coefficients were calculated separately for each of these five domains and these coefficients entered as independent variables in the ICT analyses.

The analyses were performed using the exhaustive Chi-squared Automatic Interaction Detector (CHAID) algorithm available in PASW version 21 (see Tonkin et al., 2012, for a more detailed description). The parameters for these analyses were as follows: tree depth was equal to 3; the minimum number of crime pairs allowed in parent and child nodes was 100 pairs and 50 pairs, respectively; the criterion for splitting nodes was $p < .05$ using the likelihood ratio; the number of intervals was set to 10; and a 10-fold cross-validation procedure was utilized because it is not possible to perform a leave-one-out cross-validation when running classification tree analysis in PASW.

Following the criteria established by Steadman et al. (2000) and Monahan et al. (2000), which were subsequently used by Tonkin et al. (2012) in their study of crime linkage, nodes containing less than twice, but more than half, the base rate prevalence of linked pairs were deemed to be unclassifiable. These unclassifiable cases were separated from those that were successfully classified and a further CHAID analysis run on the unclassifiable cases. This iterative process was repeated until no further cases could be classified. The classification tree analysis thus became an iterative classification tree analysis because multiple classification trees were used to generate crime linkage predictions rather than predictions coming from a single tree. Research has suggested that adopting such an iterative approach yields favorable classification results compared to 'standard', single-tree classification tree analysis (e.g., Monahan et al., 2000; Steadman et al., 2000). The same parameters described above were used for all iterations of the classification tree analysis.

Bayesian analysis.

The form of Bayesian analysis used in the current study was based on the Bayesian crime linking method developed by Salo et al. (2013). That method was originally designed to predict series membership (i.e., how likely is it that crime X belongs to series Y?). Logistic regression and classification tree analysis, however, cannot be used to predict series membership because there are typically more than two crime series within a given dataset and these procedures can only make predictions for binary outcomes. To ensure greater comparability between statistical methods, Salo et al.'s (2013) Bayesian crime linking method was adapted by the third author so that it could generate predictions for crime pairs (i.e., how likely is it that these two crimes are committed by the same person?).

In developing this method, one of the decisions that had to be made was how to quantify behavioral consistency. Three different methods are reported in the current paper:

1) Method 1

A new binary variable was created for each of the 166 offender behaviors, which took a value of 1 if the behavior was either present in both crimes in the pair or absent in both crimes and 0 otherwise. Thus, there were two ways in which an offender could demonstrate behavioral consistency in Method 1: s/he could either display the same behavior across two offenses (referred to as joint presence) or s/he could *not* display that behavior across two offenses (referred to as joint absence).

2) Method 2

A new binary variable was created for each of the 166 offender behaviors, which took a value of 1 if the behavior was present in both crimes in the pair and 0 otherwise. In this method, joint presence was the only way an offender could demonstrate behavioral consistency. This method, therefore, most closely resembled how behavioral consistency was quantified using Jaccard's coefficient in the logistic regression and ICT analyses.

3) Method 3

Method 3 created a new categorical variable for each of the 166 offender behaviors, which took a value of 1 if the behavior was present in both crimes in the pair, a value of 2 if the behavior was absent in both crimes, and 0 otherwise. In Method 3 there were two ways in which an offender could demonstrate behavioral consistency (joint presence and joint absence), but unlike Method 1, where joint presence and absence were treated as equivalent, Method 3 treated joint presence and joint absence as different types of offender behavioral consistency.

For each of these three methods, Bayesian analysis was used to model the probability of observing behavioral consistency across each of the 166 variables for both linked and unlinked crime pairs. Essentially this involved building up a picture of what the 'typical' linked pair looked like and what the 'typical' unlinked pair looked like in terms of the presence/absence of the 166 variables. A predicted probability value could then be computed, ranging from 0 (indicating that the crime pair in question was a very close fit to the 'typical picture' of an unlinked pair) up to a value of 1.00

(indicating that the crime pair was a very close fit to the 'typical picture' of a linked pair). A LOOCV method of cross-validation was used for the Bayesian analyses.

The predicted probability values produced by the logistic regression, ICT, and Bayesian analyses (ranging from 0 to 1.00) were used to construct ROC curves, which gave an indication of discrimination accuracy via the Area Under the Curve (AUC). The AUC typically ranges from 0.50 (indicating that the use of offender behavior to distinguish between linked and unlinked crime pairs is no better than chance) up to 1.00 (indicating perfect discrimination accuracy). An AUC value was calculated for each statistical method and these values compared statistically, thereby indicating the relative ability of logistic regression, ICT analysis, and Bayesian analysis to discriminate between linked and unlinked crime pairs. This allowed us to test which statistical approach offered the greatest potential for supporting the future development of crime linkage decision-support tools.

ROC analysis has been criticized, however, as a measure of discrimination accuracy in so-called 'low base rate' scenarios, where the number of positive cases is far outweighed by the number of negative cases (see Longadge, Dongre, & Malik, 2013, for a general discussion of the class imbalance problem). In such scenarios, it is possible to achieve high AUC values whilst also making a considerable number of predictive errors (false alarms in particular). Given that there are just 4,569 linked crime pairs compared to 5,651,997 unlinked pairs in the current study, this scenario can be classed as 'low base rate'. Although, it should be noted that such imbalances are not unique to crime linkage and exist in many other classification domains (e.g., risk prediction in psychiatry, the diagnosis of rare diseases, etc.).

It was, therefore, important in the current study to examine the number of decision errors associated with the AUCs obtained by the statistical methods under examination. There are a variety of ways that this could have been done (e.g., see Bennell, 2002, for a review). However, we opted to determine, for each statistical method, the frequency and proportion of decision outcomes made when restricting the proportion of false alarms to 15%. The decision outcomes we examined were: hits, misses, false alarms, and correct rejections [9]. While a 15% false alarm rate is somewhat arbitrary, this cut-off was based on discussions with linkage practitioners, who indicated to us that, when dealing with a low base-rate event in a real-world context, an attempt would need to be made to minimize the false alarm rate in order to effectively manage resources (of course, in practice, the false alarm rate could be set at any value depending upon the circumstances).

Ethics

University ethical approval for this research was granted, as was management approval from each of the five law enforcement agencies that provided data. Data were anonymized prior to being shared with the research team and were stored throughout the project on encrypted memory sticks and laptops.

III RESULTS

Three types of analysis were used to distinguish between linked and unlinked crime pairs (binary logistic regression, ICT, and Bayesian analysis), and their ability to do so was compared using ROC analysis (see Table 1 for a summary of these findings). All statistical approaches demonstrated statistically significant levels of discrimination accuracy ($p < .001$).

When the AUC values for each statistical method were compared with each other (i.e., binary logistic regression compared to ICT, binary logistic regression compared to Bayesian method 1, and so on) using the method of DeLong, DeLong, and Clarke-Pearson (1988) all 10 comparisons were statistically significant ($p < .001$). The lowest AUC was achieved by Bayesian coding method 2 (which was statistically smaller than all other approaches) and the largest AUC achieved by Bayesian coding method 3 (which was statistically larger than all other approaches). It is also worth noting that the AUC for Bayesian coding method 1 was statistically larger than both the regression and ICT models. Also, the AUC for the ICT model was statistically larger than the regression model.

As noted previously, it is possible to achieve high AUC values despite a considerable number of predictive errors (Longadge et al., 2013). A decision threshold was, therefore, adopted that capped the false alarm rate at 15%, thereby allowing us to calculate the number (and proportion) of hits, misses, false alarms, and correct rejections that occur when predicting linkage status using logistic regression, ICT, and Bayesian models. The findings in Table 2 indicate that the worst performing statistical model was Bayesian model 2 (with a 72% hit rate and a 28% miss rate) and the best performing model was Bayesian model 3 (with an 83% hit rate and a 17% miss rate).

IV DISCUSSION

Crime linkage is a considerable challenge for criminal justice practitioners (Santtila et al., 2004; Woodhams, Bull et al., 2007). One potential method for addressing these challenges is to develop computerized decision-support tools, which may help to reduce cognitive load, help analysts to select the most appropriate behaviors for linking crimes, and which can increase analytical efficiency. But, before this can be attempted, it is important to identify which statistical methods have the greatest potential for supporting the development of these tools. Using the largest, most diverse, and most ecologically valid dataset ever collected to

[9] A hit occurs when the statistical methods (logistic regression, ICT, and Bayesian analysis) predict that the two crimes in a crime pair were committed by the same person and this is true. A miss occurs when the statistical methods predict that the two crimes in a pair were committed by different persons but they were in fact committed by the same person. A false alarm occurs when the statistical methods predict that the two crimes in a pair were committed by the same person but they were actually committed by different persons. A correct rejection occurs when the statistical methods predict that the two crimes in a pair were committed by different persons and this is true.

investigate crime linkage with sexual offenses, the current study addressed this question by comparing binary logistic regression, ICT, and Bayesian analysis in terms of their ability to distinguish between linked and unlinked sexual crimes. This was the first time such comparisons had been made for this crime type.

All statistical approaches tested were able to achieve statistically significant levels of discrimination accuracy (AUCs > 0.86, p < .001). These findings, therefore, provide support for the assumptions of behavioral consistency and distinctiveness that underpin crime linkage. Moreover, they support the notion that statistical tools might be developed in the future to support the behavioral linking of sexual offenses.

In terms of which statistical approach appears to offer the greatest potential for supporting the development of such tools, the highest level of discrimination accuracy in this study was achieved by Bayesian coding method 3. This method treated the joint presence and joint absence of behavior in a crime pair as distinctly different types of offender behavioral consistency. Interestingly, a significantly lower level of accuracy was achieved when either joint presence only contributed to consistency scores (Bayesian coding method 2) or when joint presence and joint absence were treated as the same type of behavioral consistency (Bayesian coding method 1). These findings suggest that (at least sometimes) the behaviors not displayed at a crime scene can be just as important as those that are displayed by an offender. Thus, the absence of certain behaviors seems to represent a meaningful aspect of offender crime scene behavior that should be considered by both crime analysts and statistical methods when conducting crime linkage. This conclusion should not, however, be interpreted as support for using similarity coefficients (such as the Simple Matching coefficient) that incorporate joint absence in their calculations of behavioral similarity. This is because such coefficients do not treat joint presence and joint absence as distinctly different types of behavioral consistency. Thus, statistical methods that combine joint presence and joint absence (rather than treating them as distinct forms of behavioral consistency) appear to lose important information that is useful when attempting to link crimes.

TABLE 1 Receiver Operating Characteristic (ROC) Analyses Comparing Different Statistical Approaches to Crime Linkage

Statistical Approach	AUC (*SE*)	95% Confidence Interval
Binary logistic regression analysis	0.87 (0.003)	0.87 – 0.88
Iterative classification tree analysis	0.88 (0.003)	0.87 – 0.88
Bayesian analysis (Method 1)	0.89 (0.003)	0.88 – 0.89
Bayesian analysis (Method 2)	0.86 (0.003)	0.85 – 0.86
Bayesian analysis (Method 3)	0.91 (0.003)	0.91 – 0.92

Note. All AUC values $p < .001$

TABLE 2 The Frequency (and Percentage) of Classification Decisions When Using Three Statistical Approaches to Crime Linkage

		Actual Linkage Status	
		Linked	***Unlinked***
Predicted Linkage Status	***Linked***	3,427 (75%) (LR)	847,800 (15%) (LR)
		3,472 (76%) (ICT)	847,800 (15%) (ICT)
		3,518 (77%) (BA1)	847,800 (15%) (BA1)
		3,290 (72%) (BA2)	847,800 (15%) (BA2)
		3,792 (83%) (BA3)	847,800 (15%) (BA3)
	Unlinked	1,142 (25%) (LR)	4,804,197 (85%) (LR)
		1,097 (24%) (ICT)	4,804,197 (85%) (ICT)
		1,051 (23%) (BA1)	4,804,197 (85%) (BA1)
		1,279 (28%) (BA2)	4,804,197 (85%) (BA2)
		777 (17%) (BA3)	4,804,197 (85%) (BA3)

Note. The figures in Table 2 are based on a decision threshold that caps the false alarm rate at 15% (which necessarily means that the correct rejection rate is also fixed, at 85%). LR = binary logistic regression analysis; ICT = iterative classification tree analysis; BA1, BA2, BA3 = Bayesian analysis coding methods 1, 2, and 3. The top left of the table indicates the proportion of hits achieved by each method, the top right indicates the proportion of false alarms, the bottom left indicates misses and the bottom right indicates correct rejections. It should also be noted that the actual number of linked crime pairs in the sample was 4,569 and there were 5,651,997 unlinked pairs.

It was also found in this study that the ICT model marginally out-performed the logistic regression model in terms of discrimination accuracy. These findings suggest that the ability to capture/utilize differences between offenders in how they display behavioral consistency is of benefit when using statistical methods to link sexual offenses. But, the superior accuracy achieved by two out of the three Bayesian models (compared to regression and ICT) suggests that statistical approaches that utilize individual offender behaviors when generating crime linkage predictions may be preferable to approaches that rely on a single summary score of behavioral similarity.

Having discussed differences between the various statistical approaches tested in this study, it is important to note that, while the AUC values reported in Table 1 differ at a level that is statistically significant, we must be cautious not to over-estimate the practical importance of these findings. As explained by Sullivan and Feinn (2012, pp. 279-280), "[w]ith a sufficiently large sample, a statistical test will almost always demonstrate a significant difference". Given the large sample examined here, it would, therefore, seem premature to conclusively recommend one statistical approach over another until more extensive testing of these methods is conducted in real-world settings. There also needs to be consultation with crime linkage practitioners as to which method is preferable for their uses (e.g., which is the most user-friendly, produces the most interpretable output, etc.) and which statistical approaches best replicate the types of analytical task they face and the types of data they use to link crimes. Indeed, the statistical methods tested in the current study would lead to very different types of output for crime analysts, and we do not yet know the most useful way of presenting this information to analysts to support their decision-making. Examining this issue should be an aim of future research. Given the uniformly high AUC values in this study, we would suggest that all statistical approaches investigated here should be incorporated into prototype crime linkage decision-support tools in the future. Further comparisons between the methods can then be made in terms of accuracy, usability, and suitability for practice, all of which are equally important issues as researchers seek to develop computerized crime linkage tools.

An important aim of the analysis reported in this study was to estimate how many predictive errors might be expected when using these statistical approaches to link crimes (see Table 2). Our findings indicate that- despite high AUC values- a large number of predictive errors should still be expected when using statistical approaches to support crime linkage (at least in cases where crime samples are characterized by low base rates of linked crimes). But, this does not necessarily preclude the development of decision-support tools. One key question is whether the degree of decision error associated with statistical approaches is less than that associated with the existing methods used by analytical units to link crimes. A second key question is whether the level of error associated with statistical approaches is acceptable to policy-makers and senior managers who must make decisions about the overall financial and human cost of using different policing procedures. Gaining answers to these questions is vital (and should be a priority for future research), as this will help to determine whether computerized decision-support tools are truly able to support the crime linkage work of criminal justice practitioners.

In terms of future research directions and potential practical applications, the real-world testing of the findings reported in this study should be considered a priority. This would require the development of software that incorporates logistic regression, classification tree, and Bayesian-based analytical functions. To use such software, the analyst would select certain parameters (e.g., they would choose the behaviors they want to include in the analysis and specify any temporal or geographical restrictions they want to use). They would then run the analysis and the computerized crime linkage support tool would extract crime scene information from criminal justice/police databases and run this information through the statistical algorithms tested in this study (i.e. the analyst would not need to perform any complex calculations themselves, the tool would do this automatically for them). The tool would then provide the criminal justice practitioner with a list of crime pairs, ranked in terms of how behaviorally similar they are (with the most behaviorally similar at the top of the list). This would: (i) allow an analyst to analyze vast quantities of crime scene information from multiple offenses far more quickly than they would be able to manually; and (ii) would provide analysts with an evidence-based approach to prioritizing their workload.

The practical value of such tools might then be tested in a number of ways, including prospective testing where the tool is used to make predictions for unsolved crimes which are followed up over time to determine the accuracy of these predictions. Another approach would be to conduct experimental studies that require practitioners to complete mock crime linkage tasks. When completing these tasks, some practitioners would be given access to the decision-support tool to assist them, whereas other practitioners would not. The decision-making performance of these two groups would then be compared to determine whether having access to the tool conferred an advantage when completing the crime linkage task. Beyond such studies, it would also be important to evaluate how user-friendly the tool is and whether it provides all of the necessary analytical functions to meet the needs of crime linkage practitioners.

Despite the potential benefits that might be provided by computerized decision-support tools, it is important to point out that we are not suggesting that such tools replace human analysts[10]. Instead, we see such tools functioning in a similar way

[10] This is because the complex interactions between offender behavior and situational factors are not easily modeled using existing statistical methods (e.g.,

to structured professional judgment in the domain of risk assessment (e.g., the HCR-20). That is, the tool helps the practitioner to structure decision-making by emphasizing the use of empirically-informed linking cues and by helping analysts to prioritize their analytical work such that they focus on those crimes that have the greatest predicted likelihood of being linked. Ultimately, however, the practitioner decides what to do with the information and guidance provided by the linkage tool, and they always have the option of adding to this information, modifying it, or overriding it altogether if they feel they have adequate grounds to do so. Utilized in this way, there are a number of potential advantages that computerized linkage tools might offer criminal justice agencies. First, computerized tools can process large volumes of information in a quick and efficient manner (more quickly than a human analyst would be able to). At a time when police resources are being cut, any process that can potentially increase analytical efficiency is of significant value. Second, computerized crime linkage support tools would be based on empirical research (and could be updated as new findings emerge). The importance of evidence-based practice is recognized amongst criminal justice agencies around the world (see Sherman, 2013, for a review), thus the use of crime linkage decision-support tools would help criminal justice agencies to adhere to the principles of evidence-based practice.

An important consideration when developing computerized crime linkage support tools is their applicability across different jurisdictions (e.g., from one country to the next). Crime linkage practices may vary from one jurisdiction to the next and ideally any tool that is developed should be able to cope with such differences and still produce output that is useful for the criminal justice practitioner (regardless of their jurisdiction). Likewise, research suggests there is cultural variation in offender behavior (e.g., Woodhams & Labuschagne, 2012), so any tool must incorporate statistical algorithms that can account for such differences. Indeed, the impact of cultural variation on behavioral consistency, distinctiveness, and discrimination accuracy is not an issue that has been explored in sufficient depth. While it is beyond the scope of the current paper to explore this issue, it should be an aim for future research.

While the current study was concerned with linking sexual offenses, it is worth noting that the crime linkage assumptions have been tested with a range of crime types and similar suggestions made regarding the development of computerized crime linkage tools that could be used with burglary, car theft, arson, homicide, and robbery crimes (e.g., Bennell & Jones, 2005; Burrell et al., 2012; Ellingwood et al., 2013; Oatley et al., 2006; Santtila et al., 2008; Tonkin et al., 2012; Woodhams & Toye, 2007). Thus, once the basic infrastructure of a computerized decision support tool was developed, it would be possible to adapt and test the application of that tool in the linking of a range of crime types beyond sexual offenses (using the relevant statistical algorithms developed by previous research).

Finally, it is important to consider the limitations of the current study. While this research offered an improvement on prior studies in terms of including unsolved and apparent one-off offenses, it is unclear whether the proportion of solved to unsolved or serial to non-serial offenses in these data is representative of criminal justice databases (ideally they would be). Although, it should be noted that this limitation was unavoidable because it is impossible to calculate these ratios using real-world databases when we do not know whether unsolved crimes form part of a series or not. A further limitation is that UK-based offenses are over-represented in the current data compared to crimes from the other four countries. Consequently, the findings are necessarily biased towards the UK and may generalize less to other countries. Furthermore, while extensive efforts were made to match variables across countries, there will inevitably be some noise in the data in terms of cross-country coding variation. This noise would, however, only serve to reduce discrimination accuracy. Consequently, the statistically significant AUC values found in this study occur in spite of the noise rather than because of it. Another important limitation is that certain types of crime scene information (notably the geographical location of the offense) were not utilized when developing statistical linkage algorithms in this study. It is possible that the inclusion of such information would further increase discrimination accuracy. A final limitation is that, while considerable effort was taken to ensure the comparability of the different statistical methods tested in this study (e.g., by adapting Salo et al.'s (2013) Bayesian crime linking method), there are still differences between the methods that impact on our ability to draw comparisons. For example, the ICT analyses utilized five different behavioral domains when generating linkage predictions, whereas logistic regression used a single measure of behavioral consistency that combined all offender behaviors [11]. Finally, different methods of cross-validation were used for the logistic regression and Bayesian analysis (LOOCV) compared to the ICT analysis (10-fold cross-validation) because a LOOCV method is not available in PASW when conducting classification tree analysis.

Despite these limitations, the dataset utilized in the current study was substantially larger than those used in previous studies

Woodhams, Hollin, & Bull, 2008). For example, a computer might struggle to identify that behavioral changes across a crime series are due to situational factors rather than necessarily any change in the offender's motivations, fantasies, etc.

[11] A combined Jaccard's coefficient was used for the regression analyses, as this method that has typically been used in previous research and has been shown to lead to higher levels of accuracy than breaking behavior down into domains (e.g., Bennell, Jones, & Melnyk, 2009; Woodhams & Labuschagne, 2012).

of crime linkage with sexual offenses [12], which not only replicates the investigative reality faced by many crime linkage practitioners (who face large databases when linking crimes), but also increases the likelihood that the findings of this study can be generalized beyond the sample studied. Moreover, unlike many previous crime linkage studies, this study included both unsolved and apparent one-off crimes. Consequently, the current findings are more relevant to the real-world of criminal justice than those produced in previous studies (e.g., Bennell & Canter, 2002; Burrell et al., 2012; Santtila et al., 2005, 2008; Tonkin et al., 2008; Woodhams & Toye, 2007). The current study, therefore, represents an important contribution to the crime linkage literature, and helps to provide a more robust evidence base upon which to develop the practice of crime linkage. Nevertheless, the success of these endeavors rests on future research developing, trialing, and evaluating decision-support tools in real-world settings. This is our primary aim for the future.

REFERENCE

Adcock, A. B. (2000). Effects of cognitive load on processing and performance. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.387&rep=rep1&type=pdf

Bennell, C. (2002). Behavioural consistency and discrimination in serial burglary (Unpublished doctoral dissertation). University of Liverpool, Liverpool, UK.

Bennell, C., Bloomfield, S., Snook, B., Taylor, P., & Barnes, C. (2010). Linkage analysis in cases of serial burglary: Comparing the performance of university students, police professionals, and a logistic regression model. Psychology, Crime & Law, 16, 507-524. doi: 10.1080/10683160902971030

Bennell, C., & Canter, D. V. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. Science and Justice, 42, 153-164. 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., Goodwill, A. M., & Chinneck, A. (2015). Informing practice: Research methods in crime linkage analysis. In J. Woodhams & C. Bennell (Eds.), Crime linkage: Theory, research, and practice (pp. 337-367). Boca Raton, FL: CRC Press.

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 clues: Current levels of linking accuracy and strategies for moving forward. Journal of Investigative Psychology and Offender Profiling, 11, 29-56. doi: 10.1002/jip.1395

Burrell, A., Bull, R., & Bond, J. W. (2012). Linking personal robbery offences using offender behaviour. Journal of Investigative Psychology and Offender Profiling, 9, 201-222. doi: 10.1002/jip.1365

Canter, D., & Youngs, D. (2008). Interactive Offender Profiling System (IOPS). In S. Chainey & L. Tompson (Eds.), Crime mapping case studies: Practice and research (pp. 153-160). Chichester: Wiley.

Collins, P. I., Johnson, G. F., Choy, A., Davidson, K. T., & MacKay, R. E. (1998). Advances in violent crime analysis and law enforcement: The Canadian Violent Crime Linkage Analysis System. Journal of Government Information, 25, 277–284. doi: 10.1016/S1352-0237(98)00008-2

Davies, A. (1992). Rapist's behavior: A three-aspect model as a basis for analysis and the identification of serial crime. Forensic Science International, 55, 173-194. doi: 10.1016/0379-0738(92)90122-D

DeLisi, M., & Gatling, J. M. (2003). Who pays for a life of crime? An empirical assessment of the assorted victimisation costs posed by career criminals. Criminal Justice Studies, 16, 283-293. doi: 10.1080/0888431032000183489

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics, 44, 837-845. url: http://www.jstor.org/stable/2531595

Dodd, T., Nicholas, S., Povey, D., & Walker, A. (2004). Crime in England and Wales 2003/04 (Home Office Statistical Bulletin 10/04). London, UK: Home Office Research, Development and Statistics Directorate.

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 offences. Journal of Investigative Psychology and Offender Profiling, 10, 1-27. doi: 10.1002/jip.1364

Grubin, D., Kelly, P., & Brunsdon, C. (2001). Linking serious sexual assaults through behaviour (Home Office Research Study 215). London, UK: Home Office Research, Development and Statistics Directorate.

Häkkänen, H., Lindlöf, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in a sample of Finnish stranger rapes. Journal of Investigative Psychology and Offender Profiling, 1, 17-32. doi: 10.1002/jip.001

Longadge, R., Dongre, S. S., & Malik, L. (2013). Class imbalance problem in data mining: A review. International Journal of Computer Science and Network, 2. Retrieved from https://arxiv.org/ftp/arxiv/papers/1305/1305.1707.pdf

McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: New crime-specific estimates for policy and program evaluation. Drug and Alcohol Dependence, 108, 98-109. doi: 10.1016/j.drugalcdep.2009.12.002

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Clark Robbins, P., Mulvey, E. P., … Banks, S. (2001). Rethinking risk assessment: The MacArthur study of mental disorder and violence. Oxford, UK: Oxford University Press.

Oatley, G. C., Ewart, B. W., & Zeleznikow, J. (2006). Decision support systems for police: Lessons from the application of data mining techniques to "soft" forensic evidence. Artificial Intelligence and Law, 14, 35–100. doi: 10.1007/s10506-006-9023-z

Piquero, A. R., Farrington, D. P., & Blumstein, A. (2007). Key issues in criminal career research: New analyses of the Cambridge study in delinquent development. New York, NY: Cambridge University Press.

Porter, M. D. (2014). A statistical approach to crime linkage. Retrieved from http://arxiv.org/pdf/1410.2285v1.pdf

Rainbow, L. (2015). A practitioner's perspective: Theory, practice, and research. In J. Woodhams & C. Bennell (Eds.), Crime linkage: Theory, research, and practice (pp. 173-196). Boca Raton, FL: CRC Press.

Rentoul, L., & Appleboom, N. (1997). Understanding the psychological impact of rape and serious sexual assault of men: A literature review. Journal of Psychiatric and Mental Health Nursing, 4, 267-274. doi: 10.1046/j.1365-2850.1997.00064.x

Resick, P. A. (1993). The psychological impact of rape. Journal of Interpersonal Violence, 8, 223-255. doi: 10.1177/088626093008002005

Salo, B., Sirén, J., Corander, J., Zappalà, A., Bosco, D., Mokros, A., & Santtila, P. (2013). Using Bayes' theorem in behavioural crime linking of serial homicide. Legal and Criminological Psychology, 18, 356-370. doi: 10.1111/j.2044-8333.2011.02043.x

Santtila, P., Junkkila, J., & Sandnabba, 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., Korpela, S., & Häkkänen, H. (2004). Expertise and decision-making in the linking of car crime series. Psychology, Crime & Law, 10, 97-112. doi: 10.1080/1068316021000030559

Santtila, P., Pakkanen, T., Zappalà, A., Bosco, D., Valkama, M., & Mokros,

[12] Sample sizes have typically ranged from 43 to 244 offenses (Bennell et al., 2009; Santtila et al., 2005; Slater, Woodhams, & Hamilton-Giachritsis, 2014; Winter et al., 2013; Woodhams & Labuschagne, 2012). Although, there are a small number of studies that have tested larger samples, including Grubin et al. (2001) who tested two samples consisting of 468 and 840 sexual assaults respectively and Yokota et al. (2007) who tested a sample of 1,252 offenses.

A. (2008). Behavioural crime linking in serial homicide. Psychology, Crime & Law, 14, 245-265. doi: 10.1080/10683160701739679

Sherman, L. W. (2013). The rise of evidence-based policing: Targeting, testing, and tracking. Retrieved from http://cebcp.org/wp-content/evidence-based-policing/Sherman-TripleT.pdf

Slater, C., Woodhams, J., & Hamilton-Giachritsis, C. (2014). Can serial rapists be distinguished from one-off rapists. Behavioral Sciences & the Law, 32, 220-239. doi: 10.1002/bsl.2096

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Clark Robbins, P., Mulvey, E. P., … Banks, S. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. Law and Human Behavior, 24, 83-100. doi: 10.1023/A:1005478820425

Sullivan, G. M., & Feinn, R. (2012). Using effect size- or why the p value is not enough. Journal of Graduate Medical Education, 4, 279-282. doi: 10.4300/JGME-D-12-00156.1

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. W., & Santtila, P. (2012). A comparison of logistic regression and classification tree analysis for behavioural case linkage. Journal of Investigative Psychology and Offender Profiling, 9, 235-258. doi: 10.1002/jip.1367

Wilson, L., & Bruer, C. (2017). Violent Crime Linkage System (ViCLAS). Retrieved from http://www.rcmp-grc.gc.ca/to-ot/cpcmec-ccpede/bs-sc/viclas-salvac-eng.htm#countries

Winter, J., Lemeire, J., Megank, 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

Wolfgang, M. E., Figlio, R. M., & Sellin, T. (1972). Delinquency in a birth cohort. Chicago, IL: University of Chicago Press.

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: Humana Press Inc.

Woodhams, J., Grant, T. D., & Price, A. R. G. (2007). From marine ecology to crime analysis: Improving the detection of serial sexual offences using a taxonomic similarity measure. Journal of Investigative Psychology and Offender Profiling, 4, 17-27. doi: 10.1002/jip.55

Woodhams, J., Hollin, C. R., & Bull, R. (2007). The psychology of linking crimes: A review of the evidence. Legal and Criminological Psychology, 12, 233-249. doi: 10.1348/135532506X118631

Woodhams, J., Hollin, C., & Bull, R. (2008). Incorporating context in linking crimes: An exploratory study of situational similarity and if-then contingencies. Journal of Investigative Psychology and Offender Profiling, 5, 1-23. doi: 10.1002/jip.75

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 serious sexual assaults. Behavioral Sciences & the Law, 25, 841-856. doi: 10.1002/bsl.793

POSTSCRIPT

ACKNOWLEDGEMENTS

Funding. This work was supported by the Leverhulme Trust [IN-2012-114]. The funding agency was not involved in study design, data collection, analysis or interpretation, nor was it involved in the writing of this article or the decision to submit the article for publication.

The authors would also like to acknowledge the support of all five criminal justice agencies and colleagues that supported this project by providing data. Also, we would like to acknowledge the important assistance of Captain Elmarie Myburgh in translating victim accounts in South Africa and Kari Davies for help formatting the data.