Linking serial homicide - towards an ecologically valid application

Pakkanen, Tom; Sirén, Jukka; Zappalà, Angelo; Jern, Patrik; Bosco, Dario; Berti, Andrea; Santtila, Pekka

Published in:
Journal of Criminological Research Policy and Practice

DOI:
10.1108/JCRPP-01-2020-0018

Published: 01/01/2020

Linking serial homicide – towards an ecologically valid application

Tom Pakkanen, Jukka Sirén, Angelo Zappalà, Patrick Jern, Dario Bosco, Andrea Berti and Pekka Santtila

Abstract

Purpose

Crime linkage analysis (CLA) can be applied in the police investigation-phase to sift through a database to find behaviorally similar cases to the one under investigation and in the trial-phase to try to prove that the perpetrator of two or more offences is the same, by showing similarity and distinctiveness in the offences. Lately, research has moved towards more naturalistic settings, analyzing datasets that are as similar to actual crime databases as possible. One such step has been to include one-off offences in the datasets, but this has not yet been done with homicide. The aim of the present study was to investigate how linking accuracy of serial homicide is affected as a function of added hard-to-solve one-off offences.

Methodology

A sample (N = 117-1160) of Italian serial homicides (n = 116) and hard-to-solve one-off homicides (n = 1-1044, simulated from 45 cases) was analysed using a Bayesian approach to identify series membership, and a case by case comparison of similarity using Jaccard's coefficient. Linking accuracy was evaluated using Receiver Operating Characteristics (ROC), and by examining the sensitivity and specificity of the model.

Findings

After an initial dip in linking accuracy (as measured by the AUC), the accuracy increased as more one-offs were added to the data. While adding one-offs made it easier to identify correct series (increased sensitivity), there was an increase in false positives (decreased specificity) in the linkage decisions. When rank ordering cases according to similarity, linkage accuracy was affected negatively as a function of added non-serial cases.

Practical implications

While using a more natural data set, in terms of adding a significant portion of non-serial homicides into the mix, does introduce error into the linkage decision, the authors conclude that taken overall, the findings still support the validity of CLA in practice.

Originality

This is the first crime linkage study on homicide to investigate how linking accuracy is affected as a function of non-serial cases being introduced into the data.

Keywords: behavioral crime linking; serial homicide; one-off homicide; hard-to-solve homicide; ecological validity

Introduction

Behavioral crime linking refers to the practice of analyzing the crime scene behavior of two or more offences to determine if the offender could be the same (Woodhams & Bennell, 2014). The concept rests on two theoretical assumptions: offenders behave in a *consistent* way from one crime to another (Canter, 1995), and in a *distinctive* manner from other offenders committing similar crimes (Alison et al., 2002). A good number of studies has found support for the consistency and distinctiveness hypotheses across a range of crimes from violent interpersonal crime (robbery, rape, and homicide) to property crime (car theft, burglary, and arson) (Bennell et al., 2014). Moreover, results from these studies suggest that crime linkage analysis (CLA) can be used to successfully identify series of crimes: Bennell and colleagues (2014) summarized the findings of 19 crime linking studies, concluding that moderate levels of linking accuracy (as measured by the *AUC*) were achieved.

The benefit of successfully linking crimes is that individual crimes can be investigated and prosecuted as a whole, pooling valuable police resources (Grubin et al., 2001). Also, one advantage of linking crimes using behavior is that technical evidence (such as DNA and fingerprints) is not always available at the crime scene (ibid). Once automated computer systems for CLA are in place, the analysis itself is fast and cheap by comparison to technical evidence (Davies & Woodhams, 2019). CLA can be carried out in a number of ways, but during a criminal investigation, a common scenario of crime linkage analysis would be an investigator asking an analyst: "I'm investigating this one offence, can you find anything behaviorally similar in your database?" (e.g., Rainbow, 2014; for a recent and comprehensive review of the literature on the practice of crime linking, see Davies & Woodhams, 2019).

Over the last decade, increased attention has been given to the methodology used in linkage research, both in an attempt to compare the efficiency of different statistical methods (e.g. Tonkin et al., 2017; Winter et al., 2013), and in trying to increase the ecological validity of the research samples (e.g. Tonkin et al., 2011; Winter et al., 2013; Woodhams et al. 2019; Woodhams & Labuschagne, 2012). The latter is a critical issue for the application of CLA in the courtroom, as shown by a number of cases where expert opinions on crime linkage have been rejected as evidence (Bosco et al., 2010; Pakkanen et al., 2014; see also HMA v. Thomas Ross Young, 2013). In many cases, a key concern has been the generalizability of crime linkage research on CLA to actual cases, as the datasets used in research are often not representative of actual police databases. The risk here is that this distort the probability of two cases being linked, which in reality may be lower and this in turn may result in a false conviction for a case where CLA has been relied upon.

Previous crime linkage research has likely overestimated behavioral similarity because study samples have typically consisted of only solved and serial offences (Bennell & Canter, 2002; Woodhams et al., 2007). Scholars have attempted to estimate bias arising from studying solved offences by looking at crime linkage accuracy for unsolved (but linked by DNA) rape cases. For example, Woodhams and Labuschagne (2012) found that in cases of South African serial rape ($N = 119$), series that were linked by similarity in modus operandi (Jaccard's similarity coefficient of $M = .51$, $SD = .11$) displayed significantly more similarity in crime scene behavior than cases linked by DNA only ($M = .47$, $SD = .11$). Still, they found that linked pairs of crimes displayed a greater behavioral similarity than unlinked pairs and could be successfully differentiated even when the sample included unsolved series. It should, however, be noted that the number of unsolved crimes ($n = 14$) in the sample was low. Similarly, Woodhams and colleagues (2019) found, in a large ($N = 3,364$) international sample of rape cases, that the predictive accuracy of whether crime pairs were linked or not did not decrease significantly when unsolved series were added to the data ($AUC = 0.87$ to $AUC = 0.85$). Their sample had the same limitation, though, including a relatively small number of unsolved cases ($n = 92$). When comparing the unsolved to the solved series, predictive accuracy was significantly lower ($AUC = .79$ vs. $AUC = .86$).

The aforementioned issue of solved vs. unsolved offences is likely more pressing with cases of rape than homicide, as the overall clearance rate of homicide is significantly higher than that of rape (e.g. Aebi & Linde, 2012; Liem et al. 2018). The more critical issue in regard to ecological validity of crime linking research on serial homicide is thus the typical exclusion of apparent one-off offences ("apparent", because absolute certainty can rarely be achieved regarding whether an offender has offended only once). Traditionally research on crime linkage has looked at datasets comprising exclusively of serial offences (Bennell & Canter, 2002; Woodhams, 2008), while police databases (on which CLA is done in real life) include both serial and one-off offences. Tonkin and his colleagues (2011) were the first ones to consider this in their study of Finnish burglaries. They analyzed 508 burglaries with a 3:1 ratio of serial to one-off offences and found no significant differences in prediction accuracy (of whether a crime pair is linked or not) when adding one-offs to the sample. Lately a few more studies (Slater et al., 2015; Winter et al., 2013; Woodhams et al., 2019) have looked at datasets comprising both serial and apparent one-off rapes.

Winter and his colleagues (2013) analyzed 90 serial and 129 one-off rapes (at a 0.7:1 ratio) and found that linking accuracy using a naïve Bayesian classifier was not diminished (series only *AUC* = .84, one-offs added *AUC* = .89). Specifically, the sensitivity (as measured by Youden's index) of the model (i.e. the ability to correctly identify linked crimes) rose from .78 with only the serial rapes to .87 with added one-offs, while the specificity (i.e., the ability to correctly reject unlinked crimes) went down from .83 to .66.

In their study of serial (n = 194) and one-off (n = 50) rapes (ratio 3.9:1), Slater and her colleagues (2015) found, contrary to Winter (2013), that linking accuracy (as measured by the *AUC*) was not significantly affected as one-off offences were introduced to the data (series only *AUC* = .87; one-offs added *AUC* = .86). However, looking at optimal decision thresholds (as measured by Youden's index), the results were similar to that of Winter (2013): when adding one-offs, sensitivity (ability to correctly identify links) remained the same (.79) while specificity (ability to correctly reject unlinked crimes) worsened (.81 to .79). The authors noted that as the proportion of one-off offences in their sample was low, future research should investigate if an increased proportion of one-off offences causes decrements in linking accuracy.

Woodhams and her colleagues (2019) similarly looked at both serial and apparent one-off stranger rape cases in a large international sample (2173 serial and 1191 one-off rapes, at a 1.8:1 ratio). Much in line with Winter (2013) and Slater (2015), Woodhams found the addition of one-offs to have a negligible effect on linkage accuracy (series only $AUC = .86$; one-offs added $AUC = .85$). While the study is a step in the right direction in terms of mirroring the contents of an actual police database, the authors point out that the exact proportion of one-off offences to serial offences is unknown. As the exact proportion is nearly impossible to find since a database can both include seemingly one-off offences where the offender has been caught for only one offence, and undetected series, where the link between an offenders' offences has not been successfully identified, the authors suggest research be done on datasets where the proportion is varied. Only one study has tested how a varying proportion of one-off offences to serial offences affects linking accuracy. Haginoya (2016) developed samples from 840 serial and 630 one-off offences to do a series of comparisons between linked and unlinked crime pairs. The results indicated that the *AUC* between samples with varied proportions of one-off offences (0%, 25%, 50%, 75%) were comparable.

To date, no behavioral crime linking studies on homicide have included one-off homicides. For stranger rapes and burglary, it appears feasible that one-offs are more scarce than serial offences, whereas for homicides, one-off offences are expected to be more prevalent than serial homicides. The FBI estimates that less than one percent of the homicides committed in the US annually are committed by serial killers (Morton & Hilts, 2008). International estimates put the amount of serial homicides in similar proportions: approximately one percent in Australia (Mouzos & West, 2007) and in the UK (Wilson, 2007), and 1.6% in Sweden (Sturup, 2018). It is therefore of considerable importance to conduct studies including one-off offences in crime linking research on homicide.

As most one-off homicides are committed by people close to the victim (e.g., Fox & Levin, 1998; Kraemer *et al.*, 2004), they are commonly faster and easier to solve than serial homicides. As crime investigators are unlikely to seek profiling advice in these "easier to solve" homicides, and also prosecutors and defense lawyers are unlikely to ask for expert advice on crime linking in these cases, it does not make sense to model crime linkage advice after them. Indeed, the crime linkage database maintained by the Serious Crime Analysis Section (SCAS) in the UK, for example, populates their database in accordance with the same logic: the database excludes the easier to solve homicides where the offender and victim know each other, and there is no sexual element in the killing (HMIC & HMICPS, 2012). Hence, it would be prudent to identify hard-to-solve one-off homicides for the modelling.

Aim

We investigated how the prediction of series membership is affected when a dataset includes both serial homicides and hard-to-solve one-off homicides. As the exact proportion of serial to hard-to-solve one-off homicides is unknown, statistical analyses were carried out so that we varied the proportions of one-off homicides in the sample in an attempt to get a clearer understanding of how the added one-offs influence predicted series membership. We expected that adding hard-to-solve one-off homicides would decrease linkage accuracy as a function of the proportion of included hard-to-solve one-off homicides. The accuracy was expected to decrease because the added none-serial homicides were expected to add noise to the data, making the series less distinguishable.

Method

The present study utilized a sample of Italian homicides. Much like elsewhere, there has been a steady decline in the homicide rates in Italy. Homicides have gone down from 1.78 / 100,000 in 1995 to 0.67 / 100,000 in 2016. Since 2014, there have been fewer than 500 homicides annually, with an all-time low of 400 in 2016 (UNOCD, 2018). In 2014, three quarters (73.7%) of the homicides were solved (Italian National Institute of Statistics, 2017). A previous study (Pakkanen, 2006) that sought to identify hard-to-solve homicides amongst all (Finnish) homicides during a 10-year period, found that 7 in 100 cases could be considered hard-to-solve. The prevalence of serial homicides in Italy was assumed to lie somewhere between 1-2% of all homicides, as estimated elsewhere in the Western world (Morton & Hilts, 2008; Mouzos & West, 2007; Sturup, 2018; Wilson, 2007). Thus, based on estimates of clearance rates for homicide (Ministry of the Interior, n.d.; Italian National Institute of Statistics, 2017), homicides without an apparent motive, and homicides committed by a stranger to the victim (Italian National Institute of Statistics, 2017), our estimate is that the proportion of hard-to-solve to serial homicides in Italy is roughly 10 to 1. As this is a rough approximation and the exact proportion is unknown, the study utilized a design that varied this proportion in the analyses.

The serial homicides

The definition of a serial killer used was an offender who had killed more than one victim with at least 24 hours between the homicides, in accordance with a widely accepted definition (e.g. Adjorlolo & Chan, 2014; Morton & Hilts, 2008; Yaksic, 2015). A total of 23 individual killers were identified from

Italian newspapers, internet searches, and microfilms of journals in libraries. The offenders had committed a total of 116 homicides, between 1970 and 2001 (most of them between 1980 and 2001). For two-thirds (n = 15) of the offenders, court files were used, and for a third (n = 8), criminological literature and newspaper articles were consulted. All killers had been convicted in an Italian court of law.

Eight of the offenders committed their offences together with another offender. One offender committed one of his homicides with a co-offender and the rest by himself. The serial homicides included 25 victims who were killed together with another victim. The homicide series varied in terms of number of victims, and time period during which the series was committed. The median number of victims was 6 (M = 6.7, SD = 4.5); the most extensive series included 17 victims (n = 1) while the shortest series encompassed two victims (n = 3). The median duration of a series was three years (M = 4.83, SD = 5.11). In the temporally longest series, the time between the first and the last homicide was 16 years, while the offences in the shortest series took place within a single year. For an elaboration of the sample and series characteristics, please see Santtila and colleagues (2008).

The hard-to-solve one-off homicides

A total of 45 hard-to-solve killings were sampled for the one-off homicides. The operationalization of "hard-to-solve" was adapted from earlier studies on homicide (Pakkanen, 2006; Pakkanen et al., 2015). Cases where the offender was caught at the scene of the crime, or the police knew the offender at the onset of the police investigation were excluded. Additionally, for inclusion, the time period between when the case was first reported to the police and when the offender was either questioned for the first time as a suspect or caught, had to be at least 72 hours. Cases that fit the aforementioned criteria were identified mainly by enquiry from the Reparto Investigazioni Scientifiche di Roma (the department of scientific investigation within the Arma dei Carabinieri). The enquiry was supplemented by the identification of a few additional hard-to-solve cases from Italian news media. The actual case files that were used to code the data were collected from each local police district where the homicides had taken place.

The hard-to-solve one-off homicides had 56 individual offenders and 48 victims and were committed between the years 2001 and 2014. There were five cases in which more than one offender had killed one victim; in three cases three offenders worked together, and in two cases two offenders had committed the homicide together. Three cases had two victims. In the first of these, there were three offenders; in the second, two offenders; and in the third, a lone offender had killed two victims.

Sample

The present study utilized 116 cases of serial homicide and 45 hard-to-solve one-off homicides. The non-serial homicides were used to simulate 1-1044 hard-to-solve one-off homicides for a total $N = 117$-1160. This was done to account for the fact that the exact proportion of hard-to-solve to serial cases is unknown, and also because we wanted to see how linking accuracy was affected as a function of added hard-to-solve one-off cases. Some cases had more than one offender or more than one victim. All offender-victim pairings were coded separately (serial $n = 155$; one-off $n = 62$), but for the statistical analyses, only a single offender-victim pairing (one pairing per case) were used. This was done to reduce possible error of inflation of certain crime scene behaviors. In cases where a single offender had killed two or more victims, one victim was randomly chosen for inclusion.

Coding scheme

All cases (both serial and hard-to-solve one-off) were coded using the same coding scheme. The files from which the data were coded were extensive, often containing hundreds of pages. For the most part they included a summary of the findings of the pathologist, witness statements, and interrogations with the suspect (see e.g., Salfati, 1998; Pakkanen, 2006; Santtila et al., 2008). The coded variables included offence-related information (such as where the body was found, what kind of injuries the victim had, what weapon, if any, was used, etc.), socio-demographic information on the victim and offender, and situational variables (e.g. what time of day the killing occurred). Additionally, other general information, such as the age of the offender and victim, was included. Most of the variables were dichotomous and coded as either present (1) or absent (0). For most of the serial homicides court transcripts were available, and for the rest of the cases major news outlets were consulted to confirm a guilty verdict in the case. Also, the one-off offenders had all received a guilty verdict for their homicide.

For the analyses of the present study, only dichotomous offence-related variables and victim related variables ($N = 89$) were used (see Appendix 1 for a listing of the variables and their definitions). Variables with no variation (i.e., no present observations or all present observations) were eliminated, as they would not have contributed to the statistical analyses. The remaining variables had missing values ranging from 0 to 47.8 % (*Md* = 5.6 %). For the main analysis, the missing values were left as such, as the method is able to coherently handle missing information and utilize it in the modelling. For the additional analysis, (the case wise comparison using Jaccard's coefficient) the missing values were substituted with a zero.

Inter-rater reliability

The inter-rater reliability of the coding scheme was estimated using Cohen's κ (Brennan & Hays, 1992; Cohen, 1960), with a mean κ of 0.72 (*SD* = 0.13; *Minimum* = 0.59, *Maximum* = 0.89), which is generally considered to be good (e.g., Cicchetti, 1994). For the serial homicides, the data coding was conducted by two research assistants under the supervision of a senior researcher (the third author) with prior experience of the coding scheme. To clarify the coding scheme (and improve inter-rater reliability), the first and third authors discussed and refined the definitions of the variables at length before and during the coding process. For the one-off homicides, the aforementioned senior researcher (the third author) coded all of the data.

The simulation

The main reason for simulating hard-to-solve one-off homicides was pragmatic: it would have been prohibitively laborious and time-consuming to gather and code a sample of 1000+ hard-to-solve homicides, as a natural database (such as a national police register) was not available. In addition, the simulation also made it easy to vary the proportion of serial to one-off offences, which was necessary, as the exact ratio of serial to one-off offences is unknown.

The simulation was carried out utilizing all of the coded one-off homicides (*N* = 45). Each new case was created one variable at a time; for each variable, one of the original 45 cases was randomly picked and the value for that particular variable [in the original case] was chosen. This way the expected frequency of the variables in the sample of simulated one-offs matched the original one-off cases. We considered 100 distinct values for the number of simulated one-offs, with the values chosen from an equally spaced grid on logarithmic scale between 1 and 1044. The total number of samples *N* varied therefore between 117 and 1160, and the ratio of serial to one-off between 116:1 and 1:10. To account for the randomness of the simulations, 100 replicate simulated samples were created for each number of simulated one-offs. In total, we created 10,000 datasets with simulated one-offs and original serials for the analyses.

To assess the validity of the simulation, we analyzed subsets of the original one-off cases. For each number of simulated one-offs between 1 and 45, we created 100 random subsets of the original one-offs. The created datasets with the subset of original one-offs and all serials (*N* between 117 and 161) were then analyzed similarly as the datasets with simulated one-offs. Comparison of the results from the analyses of datasets with simulated and original one-offs were used as a measure of validity for the simulation. The simulation was deemed valid, if datasets with simulated one-offs yielded similar results as datasets with original one-offs.

Analyses

For the main analysis, the Bayesian crime linking method developed by Salo and his colleagues (2013) was utilized. The method is based on modeling series-specific probabilities of presence for the variables, and using Bayes theorem to turn these into probabilistic predictions that a given homicide is part of a series of homicides in the data. One of the main advantages of the method developed by Salo and colleagues over traditional statistical methods for linking crimes is that it models each series' characteristics separately (for an in depth description, see Salo et al., 2013), instead of trying to look for global patterns over the whole data, which in data this heterogeneous are very hard to find. A central disadvantage of this particular method is that it assumes an ideal dataset in which the linkage status of each case is known. In reality this is not the case: in a typical police database, only a small portion of the cases are identified as being linked to one another (i.e., the same perpetrator).

For the "learning" phase of the modelling, each homicide series was modelled separately, using the dichotomous variables (*N* = 89). The one-off homicides were considered as a class of their own (i.e., treated as their own series in the analysis). This assumes that the one-off offences are more similar to each other (in terms of the crime scene behavior and victim characteristics) than they are with the serial offences, an assumption that has some support from Pakkanen and colleagues (2015) finding that there is a qualitative difference between serial homicides and hard-to-solve one-off homicides. This difference is, further, quantifiable and large enough to allow reliable differentiation: any given hard-to-solve one-off homicide and serial homicide could be identified as such with a good accuracy (*AUC* =

.88) (Pakkanen et al., 2015). Next, the probability for each case to belong to every series was calculated separately using a leave-one-out method of cross-validation. The highest probability of series membership for each case was considered the best assessment of which series each case belonged to. To get a better sense of the predictive accuracy of the model, and for ease of comparison to earlier studies (e.g. Slater et al., 2015; Winter et al., 2013; Woodhams et al., 2019), receiver operating characteristics (ROC) analyses were carried out. In these analyses, we plotted the predicted probability against accuracy (i.e., whether the prediction was correct or not). The resulting areas under the curve (*AUC*) were used to determine how well the model fared with the addition of an increasing amount (from 1 to 1044) of hard-to-solve one-off homicides, thus linearly modeling how linking accuracy is affected when one-off homicides are added to the mix in increasing proportions. To find the optimal thresholds for sensitivity and specificity of the ROC curves (i.e., to maximize the amount of correct links and correct rejections), Youden's index was calculated, in line with previous research (e.g. Bennell & Jones, 2005; Slater et al., 2015; Tonkin et al., 2011; Winter et al., 2013).

Finally, a secondary analysis was carried out to further the understanding of the practical implications of the main result (i.e. what happens to linking accuracy when one-offs are added to the mix?), and also to counter the problem (outlined above) of assumed perfect knowledge of series membership in the data. A method developed by Craig Bennell (2002) (and subsequently used in several crime linking studies, e.g., Bennell & Canter, 2002; Burrell et al., 2012; Tonkin et al., 2017; Woodhams & Labuschagne, 2012) was used to measure the similarity, in terms of offender crime scene behavior, between all of the individual homicides. This was done by calculating Jaccard's coefficient of similarity for each pair of crimes; J = a / (a + b + c), where a is the number of behaviors present in both crimes in the pair, b the number of behaviors present in crime one but not in crime two, and c the number of behaviors absent in crime one but present in crime two. A coefficient of 0 would thus indicate no similarity whatsoever between the two cases, while 1 would indicate perfect similarity, that is, that all the same behaviors are present in both offences. Employing a leave-one-out principle, each homicide was then compared to a ranked list (from most similar to least similar) of all other homicides with and without the addition of the hard-to-solve one-off homicides. That is, in a situation where a homicide investigator would be investigating a particular case, and would ask the crime analyst for the most similar cases in the database, the main research question could be formulated as "How far up in the ranking is the first correctly linked offence with and without the one-offs added to the database?" The simulation and all the analyses were carried out using Matlab.[2]

Results

One way to inspect the validity of the simulated one-off data is to compare the results of the simulated cases to the original (n = 45) one-offs in the results of the linkage analysis. The changes in the *AUC* as a function of both the added simulated, and real one-offs are very similar: the curves (and their confidence intervals) mostly overlap (Figure 1). In other words, the *AUC*s produced by the simulated and the original hard-to-solve one-off offences did not differ from each other significantly.

The *AUC* typically ranges from 0.50 (indicating that the model is no better than chance at identifying which series an offence belongs to) up to 1.00 (indicating that the model predicts series membership perfectly). The starting point for the *AUC* – prediction accuracy when the data consisted of only serial homicides – was .88 (95% *CI* = .81–.93). When adding hard-to-solve one-off offences to the mix, there was an initial dip in the *AUC* (with 30 one-off offences added *AUC* = .85; 95% *CI* = .76–.91), after which the *AUC* steadily increased all the way to *AUC* = .90 (95% *CI* = .81–.94) with the maximum of 1044 added one-offs (N = 1160) (see Figure 1). According to commonly used criteria for interpreting the area under the curve (Swets, 1988), all the *AUC*s represent a moderate to high level of accuracy.

One possible explanation for the initial dip in the *AUC* when one-offs are added, is that with a low number of one-offs (approximately as many as the mean number of offences in a series), the model fails to distinguish the one-offs from other series (or as a series of their own), thus decreasing predictive accuracy of the model. When the number of added one-off offences grow, it becomes increasingly easier for the model to distinguish them as a separate class, and the predictive accuracy rises respectively.

[2] The code is freely available at https://github.com/jpsiren/CL1off

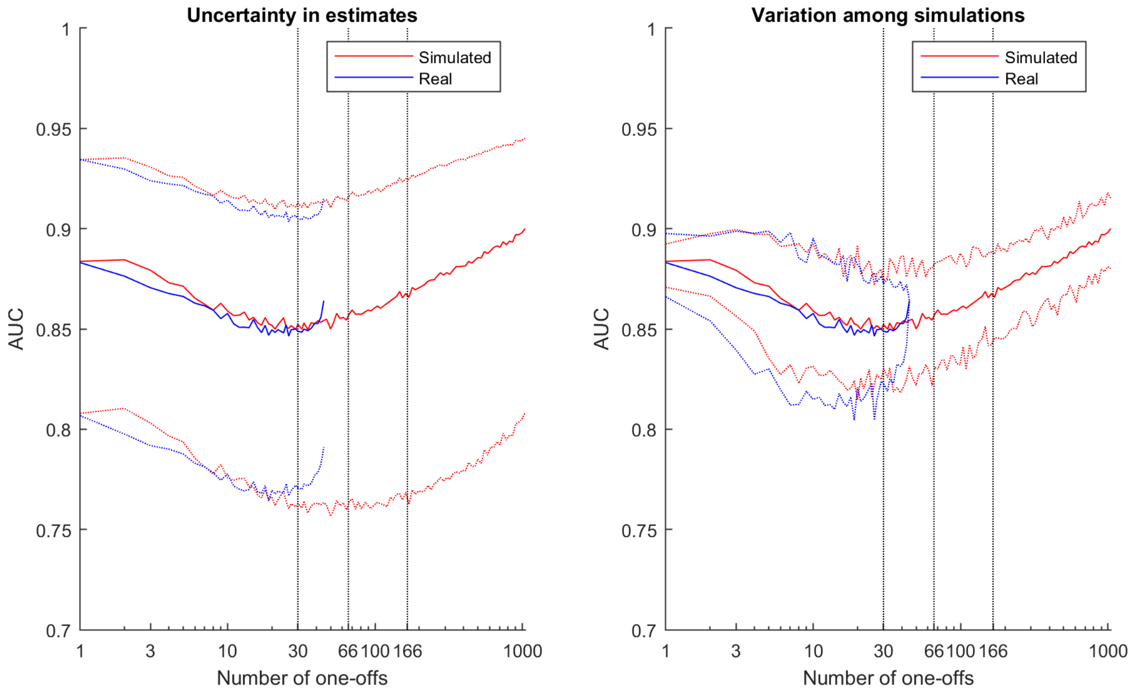

Figure 1. The area under the curve (*AUC*) as a function of added hard-to-solve one-off homicides (*n* = 1-1044; *N* = 117-1160). The 95% confidence intervals are for the uncertainty in the estimates (right) and for the variation among the simulations (left).

Note. The scale on the x-axis is logarithmic for ease of interpretation. The dotted vertical lines are for comparison to Slater et al. (2015) at 30 one-offs, Woodhams et al. (2019) at 66 one-offs, and Winter et al. (2013) at 166 one-offs (i.e., marker-lines at the same ratio of serial to one-off offences as in their samples).

Next the sensitivity and specificity were examined as a function of added one-offs. Youden's index was calculated to find the optimal decision threshold, maximizing sensitivity (identifying correct series) and specificity (rejecting false series). There is an initial decrease in sensitivity (from .76 to .70), as one-offs are added to the data, most likely because the model cannot yet identify the specific characteristics of the one-offs. This means that some of the added one-offs are mistakenly linked to different series, and possible some divergent serial cases are erroneously linked to the one-off category. This is also reflected by the larger margins of error (see Figure 2). When the number of one-offs increases, sensitivity increases substantially (to .89), and similarly, the uncertainty becomes smaller. The specificity decreases steadily with the added first 100 one-off cases (from .92 to .80), after which it starts increasing some (to .84). In other words, as a function of the added one-offs it becomes easier to identify series correctly (i.e. correct positives increase), but at the same time a larger number of series are identified erroneously (i.e. also false positives increase) (Figure 2).

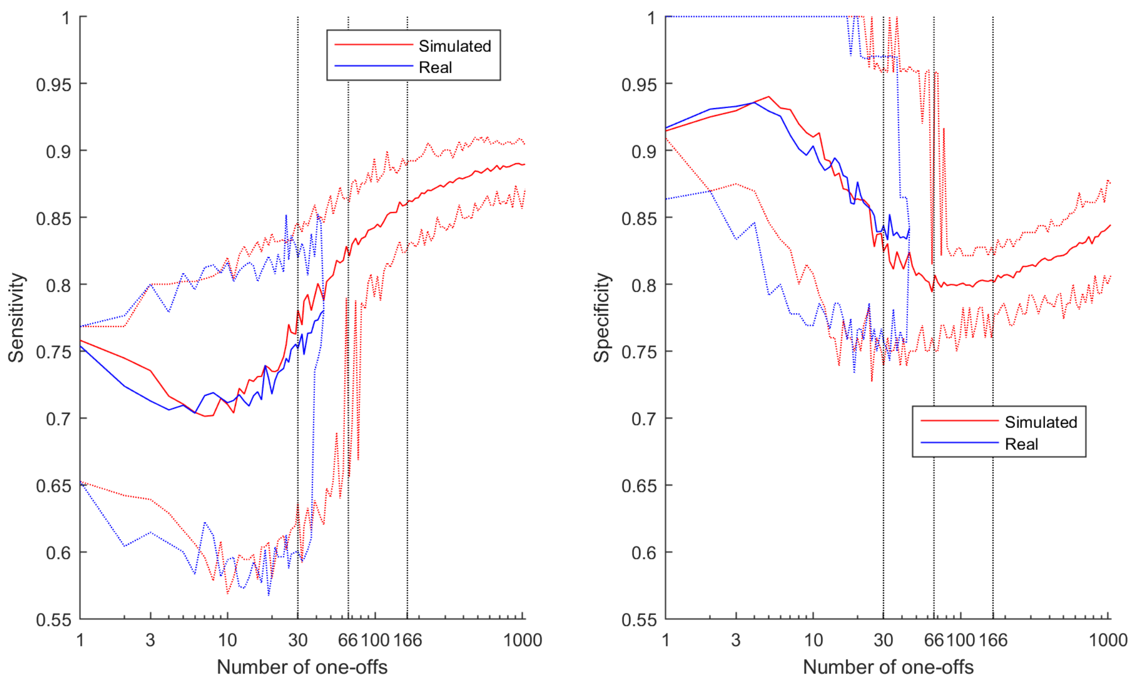

Figure 2. Sensitivity (left) and specificity (right) as a function of added one-offs, using Youden's index as a cut-off. The 95% confidence intervals are for the variation among the simulations.

Note. Youden's J = pH + pCR –1, where pH is the probability of a hit and pCR is the probability of a correct rejection (Bennell and Jones, 2005). The scale on the x-axis is logarithmic for ease of interpretation. The dotted vertical lines are for comparison to Slater et al. (2015) at 30 one-offs, Woodhams et al. (2019) at 66 one-offs, and Winter et al. (2013) at 166 one-offs (i.e., marker-lines at the same ratio of serial to one-off offences as in their samples).

In the last stage of the analysis, the offences were compared case-by-case, using Jaccard's similarity coefficient. One by one the cases were compared to a ranking of all the other cases, from most similar to least similar to the query case. A linked offence with the rank 1 would be a perfect result (a "hit"): the most similar case in the data set (as measured here by Jaccard's similarity coefficient) is linked to the index offence (Yokota & Watanabe, 2002). The hit rate in this study was 85.3% with no one-offs in the data, and 82.8% with all the 1044 one-offs added. In other words, for over 80% of the cases, the most similar case could be found at the top of the ranking list, regardless if the data included one-offs or not.

Since this does not take into account the varying length of the different series, an additional comparison was made, where a corrected median rank of each series was used, and the proportion of series where the median rank was among the 5 most similar cases, and 20 most similar cases were calculated. With the median rank correction, 56.0% of the cases could be found among the 5 most similar, when no one-offs were in the data. This dropped down to 50.7% with all the one-offs added. Similarly, 76.5% of the cases could be found among the 20 most similar cases with no-offs included, and a respective 63.2% with all the one-offs included (Figure 3).

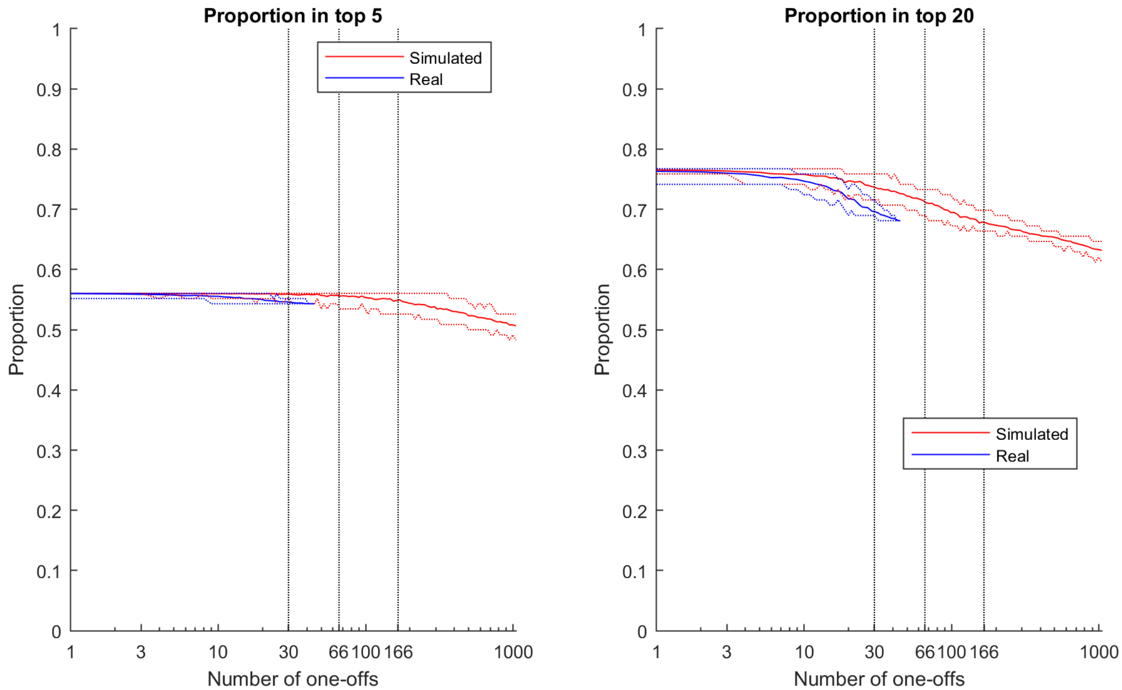

Figure 3. Case by case comparison of similarity using Jaccard's coefficient: the proportion of series with their median rank among the top 5 most similar cases (left) and the top 20 most similar cases (right) as a function of added hard-to-solve one-off homicides (n = 1-1044; N = 117-1160).

Note. The scale on the x-axis is logarithmic for ease of interpretation. The accuracy for the simulated cases is an average of 100 simulations. The dotted vertical lines are for comparison to Slater et al. (2015) at 30 one-offs, Woodhams et al. (2019) at 66 one-offs, and Winter et al. (2013) at 166 one-offs (i.e., marker-lines at the same ratio of serial to one-off offences as in their samples).

Discussion

Linking accuracy, as measured by the *AUC*, when hard-to-solve one-off homicides were added to the data, decreased initially, and started increasing steadily again, when the one-offs were more than 30. If compared with the same proportion of serial cases to one-off cases in the data, the present findings are most like Woodhams and colleagues (2019) and Slater and colleagues (2015) with a slight decrease in linking accuracy (Woodhams et al. .86 to .85; Slater et al. .87 to .86; present study .88 to .86 and .88 to .85 respectively). With the same proportion of serial cases to one-off cases as Winter and colleagues (2013), the present study had a slight decrease in accuracy (.88 to .87), while Winter and colleagues saw an increase in accuracy (.84 to .89). When we increased the number of one-offs significantly, to a proportion more realistic in terms of what a national police database might look like (10 one-offs: 1 serial), there was an increase in linking accuracy (.88 to .90).

When examining sensitivity and specificity of the linkage predictions (setting cut-offs using Youden's index), the findings of the present study are in line with Winter and colleagues (2013) and Slater and colleagues (2015). While identifying correct links becomes easier (sensitivity increases), rejecting false links becomes harder (specificity decreases), introducing false positive errors into the linkage decisions. In other words, as more hard-to-solve one-off homicides are added to the data, the model makes more mistakes by identifying links that are not real. This finding supports the hypothesis that the added non-serial cases increase error in the linkage decisions. For the investigation-phase, sensitivity is likely to be preferred over specificity, as the investigator would want to make sure not to miss any possible links. For the trial-phase, on the other hand, this (using Youden's index as a cut-off) might pose a bigger problem, as false positives could arguably be considered worse in a court of law, than false negatives.

Taking a closer look at the probabilities produced by the modelling in the present study; the magnitude of the probabilities is considerably smaller for the one-offs than for the serial cases. In other words, the model finds less similarity between the one-offs and any series, than within the serial homicides. This finding is line with Pakkanen and colleagues (2015) finding, that hard-to-solve one-offs are qualitatively different from serial homicides, and perhaps do not interfere with linkage accuracy, as they are identified as belonging to a separate class by the model. Research comparing serial homicides to one-off homicides (e.g. Fox & Levin, 1998; Kraemer et al., 2004; Pakkanen et al., 2015) have noted that serial killers tend to be more maladjusted and pathological, that their motives (e.g. overrepresentation of sexual motives amongst serial killers), and crime scene behavior (e.g. serial killers display a higher level of forensic awareness at the crime scene) differ from each other. The findings of the present study would seem to suggest that the differences between these two types of killers are quite distinguishable by crime linkage models.

For the cases by case comparison and ranking, the results would suggest a bigger practical (and negative) effect of adding one-offs to the data, than linking accuracy measured by the *AUC*. The drop in the proportion of cases to be found among the top most similar cases is noticeable. From a practical standpoint this would suggest that a crime investigator would have to go through more homicide cases than previously thought, in order to maximize their chances of finding linked cases in the crime database.

Conclusion

Against expectation, overall crime linking accuracy (as measured by the *AUC*) of homicides increased (after an initial decrease), as hard-to-solve one-offs were added to the data. Examining sensitivity and specificity more closely and looking at the case by case similarity comparison we found that, in line with the hypothesis, adding hard-to-solve one-offs does introduce error into the linkage decisions by increasing false positives. Taken as a whole, the effects seem manageable in scope, and do not thus invalidate the viability of crime linkage, even when the data used more closely mirrors real crime databases.

Replication is needed with other samples of serial and hard-to-solve one-off homicide, though. Future research should strive to include one-offs, because of the increased ecological validity, and because according to the present findings, their addition does affect CLA. For the practice of crime linking, we would say that the present findings take us one step closer to refining and fine-tuning automated algorithms to help sift through police databases to be used in the investigation-phase, and to establish more accurate error rates for our estimates of CLA in the trial-phase.

Limitations

A major methodological limitation of the current study is that the linkage status is known for all the cases. This is obviously not the case with real crime databases, a fact, which may lead to overestimation of linkage accuracy when predicting series membership in the present study. In order to circumvent for this limitation, the method would need to be developed so that either the “learning” phase of the modelling is done utilizing only a part of the linked series, or that series membership (or linkage status) is estimated in other ways.

The generalizability of the results of the present study, as such, to other countries and databases are likely limited. The homicides themselves, both serial and hard-to-solve one-offs, may very well have some cultural specificities (e.g. mafia related killings in Italy), that may eventually affect the results of the modelling. This would also be the case regarding the nature of the crime databases and the culture of crime investigation with the police in different countries. Thus, future research should strive to replicate the modelling in the countries and with databases where the CLA is to be applied.

Future research

To further increase the ecological validity of the data, and the research, the next natural step would be to conduct the analyses on natural crime databases. This would give an even more realistic estimate of the applicability and limitations of CLA in practice.

Another avenue that would be beneficial to explore further is the overlap of (serial) homicide with other crimes, specifically rape, but also perhaps burglary and robbery. From a behavioral standpoint, the distinction between rape and homicide (that the present study has compared), may be serendipitous; for example, the offender brought a knife along and the victim tried to flee, which turned an intended rape into a homicide. Or a robber that tries to extort the PIN-number for their victim's credit card and ends up killing the person in a crime where the initial intent was financial gain. While some comparative research like this exists, models like the one used in the present study could be tweaked and modified to be applied on, for example, naturalistic databases of other crime types to develop and extend CLA even further.

References

Aebi, M., & Linde, A. (2012). Crime trends in Western Europe according to official statistics from 1990 to 2007. In J. van Dijk, A. Tseloni, & G. Farrell (Eds.), *The international crime drop* (pp. 37-75). Palgrave Macmillan, London.

Adjorlolo, S., & Chan, H. (2014). The controversy of defining serial murder: Revisited. *Aggression and Violent Behavior, 5*, 486–491.

Alison, L., Bennell, C., Mokros, A., & Ormerod, D. (2002). The personality paradox in offender profiling: A theoretical review of the processes involved in deriving background characteristics from crime scene actions. *Psychology, Public Policy, and Law, 8*, 115-135.

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

Bennell, C., & Canter, D. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. *Science & Justice, 42*, 153-164.

Bennell, C., & Jones, N. (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.

Bennell, C., Mugford, R., Ellingwood, H., & Woodhams, J. (2014). Linking Crimes Using Behavioural Clues: Current Levels of Linking Accuracy and Strategies for Moving Forward. *J. Investig. Psych. Offender Profil., 11*, 29-56. doi:10.1002/jip.1395

Bosco, D., Zappalà, 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.

Brennan, P., & Hays, B. (1992). The kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. *Research in Nursing & Health, 15*, 153-158.

Burrell, A., Bull, R., & Bond, J. (2012). Linking personal robbery offences using offender behaviour. *Journal of Investigative psychology and offender profiling*, *9*, 201-222.

Canter, D. (1995). Psychology of offender profiling. In R. Bull & D. Carson (Eds.), *Handbook of psychology in legal contexts* (pp. 343-355). Chichester, England: Wiley.

Cicchetti, D. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment, 6*, 284-290.

Cohen, J. (1960). Coefficient of agreement for nominal scales. *Education and Psychological-Measurement, 20*, 37-46.

Davies, K., & Woodhams, J. (2019). The practice of crime linkage: A review of the literature. *Journal of Investigative Psychology and Offender Profiling, 16*, 169-200.

Fox, J., & Levin, J. (1998). Multiple homicide: patterns of serial and mass murder. *Crime and Justice, 23*, 407-55.

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.

Haginoya, S. (2016). Examining Linkage Analysis in Serial and Non-serial Crimes. *Japanese Journal of Applied Psychology, 42*, 112-120.

Her Majesty's Advocate v. Thomas Ross Young (2013), HCJAC 145, Appeal no. XC896/07, UK, available at: www.scotcourts.gov.uk/search-judgments/judgment?id=523286a6-8980-69d2-b500-ff0000d74aa7 (accessed 22 May, 2020).

Her Majesty's Inspectorate of Constabulary (HMIC) & Her Majesty's Inspectorate of the Crown Prosecution Service (HMICPS) (2012). *Forging the links: rape investigation and prosecution, HMIC and HMICPS*, London, available at: www.justiceinspectorates.gov.uk/hmicfrs/media/forging-the-links-rape-investigation-and-prosecution-20120228.pdf (accessed 28 July 2018).

Italian National Institute of Statistics (ISTAT) (2017). "Crimes, defendants and victims of crime – An integrated reading of the sources on crime and justice", National Institute of Statistics, Rome, available at: https://www.istat.it/it/files//2017/10/Delitti-imputati-e-vittime-dei-reati.pdf (accessed 9 October 2018).

Kraemer, G., Lord, W., & Heilbrun, K. (2004). Comparing single and serial homicide offenses. *Behavioral Sciences and the Law, 22*, 325-43.

Liem, M., Suonpää, K., Lehti, M., Kivivuori, J., Granath, S., Walser, S., & Killias, M. (2019). Homicide clearance in Western Europe. *European journal of criminology*, *16*, 81-101.

Ministry of the Interior (n.d.), "Statistics on the clearance rate of homicide in Italy 2004-2005", Ministry of the Interior, Rome, available at: www1.interno.gov.it/mininterno/export/sites/default/it/assets/files/8/ 2006817111512.pdf (accessed 27 July 2018).

Morton, R. and Hilts, M. (Eds.) (2008), *Serial Murder: Multi-Disciplinary Perspectives for Investigators*, Federal Bureau of Investigation, Quantico, VA.

Mouzos, J., & West, D. (2007). *An examination of serial murder in Australia*. Canberra: Australian Institute of Criminology.

Pakkanen, T. (2006). *Profiling hard-to-solve homicides – Identifying dimensions of offending and associating them with situational variables and offender characteristics* (Unpublished master's thesis). Åbo, Finland: Åbo Akademi University.

Pakkanen, T., Santtila, P., & Bosco, D. (2014). Crime linkage as expert evidence–making a case for the Daubert standard. In J. Woodhams & C. Bennell (Eds.), *Crime linkage: Theory, research, and practice* (pp. 225-250). Boca Raton, FL: CRC Press.

Pakkanen, T., Zappalà, A., Bosco, D., Berti, A., & Santtila, P. (2015). Can hard-to-solve one-off homicides be distinguished from serial homicides? Differences in offence behaviours and victim characteristics. *Journal of Criminal Psychology, 5*, 216-232.

Rainbow, L. (2014). A practitioner's perspective: Theory, practice, and research. In J. Woodhams & C. Bennell (Eds.), *Crime linkage: Theory, research, and practice* (pp. 173-194). Boca Raton, FL: CRC Press.

Salfati, G. (1998). *Homicide: a behavioural analysis of crime scene actions and associated offender characteristics* (Unpublished doctoral dissertation). Liverpool, UK: University of Liverpool.

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.

Santtila, P., Pakkanen, T., Zappala, A., Bosco, D., Valkama, M., & Mokros, A. (2008). Behavioural crime linking in serial homicide. *Psychology, Crime & Law*, *14*, 245-265.

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, 30*, 261-273.

Sturup, J. (2018) Comparing serial homicides to single homicides: A study of prevalence, offender and offence characteristics in Sweden, *Journal of Investigative Psychology and Offender Profiling*, DOI: 10.1002/jip.1500

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

Tonkin, M., Santtila, P., & Bull, R. (2011). The linking of burglary crimes using offender behaviour: Testing research cross-nationally and exploring methodology. *Legal and Criminological Psychology, 17*, 276–93. doi: 10.1111/j.2044-8333.2010.02007.

Tonkin, M., Pakkanen, T., Sirén, J., Bennell, C., Woodhams, J., Burrell, A., Imre, H., Winter, J., Lam, E., ten Brinke, G., Webb, M., Labuschagne, G., Ashmore-Hills, L., van der Kemp, J., Lipponen, S., Rainbow, L., Salfati, G., & Santtila, P. (2017) Using offender crime scene behavior to link stranger sexual assaults: A comparison of three statistical approaches. *Journal of Criminal Justice, 50*, 19-28.

Wilson, D. (2007). *Serial killers: Hunting Britons and their victims, 1960 to 2006*. Winchester: Waterside Press.

Winter, J., 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.

Woodhams, J., Hollin, C., & Bull, R. (2007). The psychology of linking crimes: a review of the evidence. *Legal and Criminological Psychology, 12*, 233-49.

Woodhams, J. (2008). *Juvenile sex offending: An investigative perspective* (Unpublished doctoral dissertation). Leicester, UK: University of Leicester.

Woodhams, J., & Labuschagne, G. (2012). A test of case linkage principles with solved and unsolved serial rapes. *Journal of Police and Criminal Psychology*, *27*, 85-89. doi:10.1002/s11896-011-9091-1.

Woodhams, J., & Bennell, C. (2014). Testing the Theories Underpinning Crime Linkage. In J. Woodhams & C. Bennell (Eds.), *Crime linkage: Theory, research, and practice* (pp. 132-165). Boca Raton, FL: CRC Press.

Woodhams, J., Tonkin, M., Burrell, A., Imre, H., Winter, J., Lam, E., ten Brinke, G., Webb, M., Labuschagne, G., Bennell, C., Ashmore-Hills, L., van der Kemp, J., Lipponen, S., Pakkanen, T., Rainbow, L., Salfati, G., & Santtila, P. (2019) Linking serial sexual offences: Moving towards an ecologically valid test of the principles of crime Linkage. *Legal and Criminological Psychology, 24*, 123-140.

Yaksic, E. (2015). Addressing the challenges and limitations of utilizing data to study serial homicide. *Crime Psychology Review, 1*, 108–134.

Yokota, K., & Watanabe, S. (2002). Computer-based retrieval of suspects using similarity of modus operandi. *International Journal of Police Science & Management, 4*, 5-15.

UNOCD (2018). *International homicide victims*. United Nations Office on Drugs and Crime, Vienna, available at: https://dataunodc.un.org/crime/intentional-homicide-victims (accessed 28.7.2018).

Appendix 1

The coding scheme: Variables used in the analyses ($N = 89$) and their definitions for being coded as present (1).

Variable	Definition
	Offence related variables
Point of fatal encounter	Point of fatal encounter (where the killer and the victim initially met) and the murder scene are different places.
Abduction	Victim forcibly removed from one place to another while still alive.
Body moved	Body was not found at the scene of the murder
Body in building	Body was found inside a building.
Body outside	Body was found outside.
Body in vehicle	Body was found inside a vehicle.
Body in water	Body was found immersed in water.
Body covered	Body was covered by something but not inside an object.
Body in bag	Body was covered by putting it inside a bag or a suitcase.
Body buried	Body was buried in the ground.
City	Area where the body was found was inner city.
Suburb	Area where the body was found was a suburb.
Countryside	Area where the body was found was in the countryside.
Uninhabited	Area where the body was found was uninhabited.
Rape	Killing occurred in association with a rape.
Robbery or burglary	Killing occurred in association with a robbery or burglary.
Other crime	Killing occurred in association with another crime.
Murder scene burned	Murder scene was burned in order to destroy the body or evidence.
Murder scene victim's home	Murder scene was the victim's home.
Disguise	Offender wore a disguise of some kind.
Blindfold	Use of any physical interference with the victim's ability to see.
Bound	Hands or legs of the victim bound during the attack.
Binding to scene	Object used in binding brought by the offender.
Binding at scene	Object used in binding found at the scene by the police.
Binding from scene	Object used in binding taken from the scene by the offender.
Gagging	Object used in prevention or noise (not manual gagging of the victim).
Gag to scene	Object used in gagging brought by the offender.
Gag from scene	Object used in gagging taken from the scene by the offender.
Single violence	Only a single act of violence was directed at the victim by the offender.
Clothed	Victim found fully clothed.
Naked	Victim found fully naked.
Partially unclothed	Victim found partially unclothed.
Genitals exposed	Victim found with genitals exposed.

Sex with victim	Any evidence of achieved or attempted vaginal penetration, oral penetration, anal penetration, or that the offender had ejaculated.
Object penetration	Victim penetrated with an object.
Necrophilia	Any evidence of postmortem sexual activity.
Picquerism.	Any evidence of picquerism.
Forensic awareness	Steps taken by the offender to ensure no evidence can be obtained.
Firearm	Handgun, shotgun or rifle used in the killing.
Touch shot	Victim shot so that the firearm has touched the body when fired.
Sharp weapon	Sharp weapon (such as a knife or an axe) used in the killing.
Multiple stab same	Several stab wounds to the same body area.
Multiple stab several	Stab wounds to several body areas.
Strangulation object	Victim strangled with an object.
Strangulation hands	Victim strangled manually.
Suffocation	Victim suffocated by other methods than strangulation.
Kick or hit	Victim was kicked or hit (without a weapon).
Multiple hit	Victim was hit (without a weapon) several times.
Blunt weapon	Blunt weapon used in the killing.
Multiple blunt	Victim hit with a blunt weapon several times.
Excessive blunt	Blunt weapon used excessively (more than needed to kill the victim).
Torture or humiliation	Victim was tortured or publicly humiliated.
Objects thrown	Objects were thrown at the victim.
Weapon to scene	Weapon brought by the offender
Weapon at scene	Weapon found by the police at the scene.
Weapon from scene	Weapon taken from the scene by the offender.
Body parts removed	Body parts removed from the victim.
Removed parts found	Removed body parts were found.
Head area	Injuries on the victim to the eyes, nose, mouth, or head
Throat	Injuries on the victim to the throat.
Torso	Injuries on the victim to the torso.
Extremities	Injuries on the victim to the hands, arms, legs, or feet.
Genitals	Injuries on the victim to the genitals.
Back	Injuries on the victim to any of the areas on the backside of the body.
	Victim related variables
Victim gender	Victim was male.
Alcohol	Victim was under the influence of alcohol during the attack.
Drugs	Victim was under the influence of drugs or medicines during the attack.
Student	Victim was a student or pupil.
Employee	Victim was an employee.
Unemployed	Victim was unemployed.
Prostitute	Victim was a prostitute.

Handicapped	Victim had a mental of physical handicap.
Health problems	Victim had physical health problems.
Foreigner	Victim was a foreigner, refugee, or immigrant.
Gay	Victim was known to have engaged in same-sex sexual behavior.
Relationship	Victim was married, had a same-sex registered relationship, or was currently in a serious relationship.
Divorced	Victim was divorced.
Children	Victim had children.
Drug habit	Victim had a drug habit.
Psychiatric medication	Victim had current or former psychiatric medication.
Homeless	Victim was without accommodation at the time of the killing.
Institution	Victim lives in an institution (hospital, youth home, prison).
Alone	Victim lives alone.
Cohabitation	Victim lived with an intimate partner, (a) parent(s), children, other relatives, or with a flat mate.
Owns apartment	Victim owns his or her apartment.
Rent	Victim lives in a rented accommodation.
Council flat	Victim lives in a rented accommodation owned by the city council.
Night shelter	Victim used night shelters.
Other’s flat	Victim is staying at someone else’s flat.