COMPUTERIZED CRIME LINKAGE SYSTEMS

A Critical Review and Research Agenda

CRAIG BENNELL
Carleton University

BRENT SNOOK
SARAH MACDONALD
Memorial University of Newfoundland

JOHN C. HOUSE
Royal Newfoundland Constabulary

PAUL J. TAYLOR
Lancaster University

Computerized crime linkage systems are meant to assist the police in determining whether crimes have been committed by the same offender. In this article, the authors assess these systems critically and identify four assumptions that affect the effectiveness of these systems. These assumptions are that (a) data in the systems can be coded reliably, (b) data in the systems are accurate, (c) violent serial offenders exhibit consistent but distinctive patterns of behavior, and (d) analysts have the ability to use the data in the systems to link crimes accurately. The authors argue that there is no compelling empirical support for any of the four assumptions, and they outline a research agenda for testing each assumption. Until evidence supporting these assumptions becomes available, the value of linkage systems will remain open to debate.

Keywords: crime linkage systems; comparative case analysis; behavioral linking; serial crimes; behavioral consistency; policing

An important task in some police investigations is to determine whether or not a set of crimes has been committed by the same offender (Grubin, Kelly, & Brunsdon, 2001). To assist with this linking task, law enforcement agencies have developed computerized crime linkage systems that contain offense, offender, and victim information, all of which are extracted from investigative files (Collins, Johnson, Choy, Davidson, & MacKay, 1998). The analysis of data contained within these systems is assumed to increase the probability of identifying a crime series. However, despite the widespread use of some linkage systems, the assumptions underlying these systems have seldom been tested empirically. The goal of this article is to examine these assumptions in light of available evidence and to propose an agenda for future research that can further refine our understanding of when such systems can be effective. However, before turning to our primary task, we outline the origins of crime linkage systems and describe some of their possible functions.

AUTHORS' NOTE: *Correspondence concerning this article should be addressed to Craig Bennell, Department of Psychology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6, Canada; e-mail : craig_bennell@carleton.ca.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 39 No. 5, May 2012 620-634
DOI: 10.1177/0093854811435210

THE ORIGINS AND FUNCTIONS OF CRIME LINKAGE SYSTEMS

Crime linkage systems can be traced back to the development of the Federal Bureau of Investigation's (FBI) Violent Criminal Apprehension Program (ViCAP) in 1985. ViCAP was developed for the laudable purpose of avoiding "linkage blindness," a term used to describe the absence of communication between law enforcement agencies that might be investigating related cases (Egger, 1984). The ViCAP initiative sought to reduce this problem by helping agencies determine whether or not linked crimes were being committed across jurisdictional boundaries (FBI, n.d.). To accomplish this goal, information about violent offenses was entered into a computer database and analyzed to identify crimes that showed distinct patterns of similarity that might reflect linkages. To this day, ViCAP remains a nationwide data information system for collecting, sorting, and analyzing solved and unsolved cases of violent crime (FBI, n.d.).

Other systems were developed subsequently to assist with linkage analysis. These include Washington State's Homicide Investigation Tracking System, New Jersey's Homicide Evaluation and Assessment Tracking System, Iowa's Sex Crimes Analysis System, the Royal Canadian Mounted Police's Major Crime File, and its successor, the Violent Crime Linkage Analysis System (ViCLAS; see Collins et al., 1998, for a list of additional systems). Although we believe that most of the issues discussed in this article apply to all linkage systems, our arguments often focus on ViCLAS because it is the most frequently used of all linkage systems and is generally considered the "gold standard" (Collins et al., 1998, p. 277). Currently, there are ViCLAS centers in every Canadian province, with the exception of Prince Edward Island, and its use is mandated in two provinces (Ontario and Quebec). ViCLAS is also reportedly used as part of a repertoire of investigative tools in the following locations: Australia, Austria, Belgium, Czech Republic, France, Germany, Ireland, the Netherlands, New Zealand, Switzerland, the United Kingdom, and two U.S. states (Indiana and Tennessee; Royal Canadian Mounted Police [RCMP], n.d.).

Linkage systems differ from each other in a variety of important ways, and the same system may even be used differently across jurisdictions (e.g., Witzig, 2003). However, there is arguably a core procedure that is *generally applicable* to most systems when linking several crimes. As an example, in ViCLAS, the general linking procedure consists of five broad steps (RCMP, n.d.). First, data related to specific crime types are collected and recorded, usually with the assistance of a coding manual.[1] Second, the data are scrutinized for the purpose of quality assurance, and attempts are made to fix any errors that were made during the data coding phase. Third, data are entered into a computer database that contains equivalent information about other crimes. Fourth, the data are examined for potential crime linkages, typically by someone who is trained to search the database. Fifth, once the search for linked crimes is complete, relevant investigators are informed about potential linkages. These investigators are encouraged to confirm or eliminate the potential linkages through further investigation.

There are a number of different functions that can potentially be served by crime linkage systems. The first and most obvious use is to conduct linkage analysis by searching the database to identify crimes that share similar, but distinctive, features. Single aspects of a crime can be used to conduct these searches, or the user can generate search queries that include complex combinations of the coded variables. One aspect of the data that is sometimes relied on for this purpose is the offender's behavior at the crime scene (Martineau &

Corey, 2008), although searches can be based on offense, offender, and/or victim information. Searches using crime scene behaviors can rely on aspects of an offender's modus operandi (MO), which is defined as a behavior or set of behaviors exhibited by the offender that allow him or her to successfully carry out the crime (Kangas, 2001; Martineau & Corey, 2008; Ressler, Burgess, & Douglas, 1988). However, because MO can change across an offender's crimes (Douglas & Munn, 1992), some analysts may also search for "behavioral signatures" (Gault, 2010; Keppel, 2000; Keppel, Weis, Brown, & Welch, 2005). Unlike MO, behavioral signatures are thought to be unnecessary for the successful completion of a crime but instead represent distinctive behaviors exhibited by an offender across his or her crimes to satisfy some psychological need (e.g., positioning a victim's body in a particularly degrading manner after death; Douglas & Munn, 1992).

Second, a number of individuals have suggested that crime linkage systems can serve other investigative functions that go beyond the specific task of linking crimes. For example, one function is appraisal, in that completing a coding booklet can provide the means to evaluate the effectiveness of an investigation. As Cooper (2007) argues with respect to ViCAP, "[T]he ViCAP form . . . serves as an excellent reference guide while conducting a thorough and well thought-out investigation. . . . If the questions posed by the ViCAP form are completed, it can be presumed that the investigation is complete and thorough." Given how comprehensive most coding booklets are, this argument likely applies to other linkage systems as well.

Third, although limited in practice, crime linkage systems could also potentially be used in court, where questions are sometimes raised about whether or not a defendant is responsible for a series of crimes (Bosco, Zappalà, & Santtila, 2010; Meyer, 2007; Ormerod & Sturman, 2005). Law enforcement personnel have provided testimony in court about the likelihood of multiple crimes being committed by the same offender on the basis of distinctive crime scene behaviors (e.g., Labuschagne, 2006; *State v. Code*, 1994; *State v. Pennell*, 1989; *State v. Prince*, 1992). However, when presenting their testimony, these individuals have typically not relied on linkage systems to generate their testimony. Several recent court cases have stressed the importance of basing such testimony on reliable databases (e.g., *State v. Fortin*, 2004). One advantage of having access to a large database of crimes is that it will allow the courts to determine, in a more precise fashion, the degree to which a crime scene behavior (or set of behaviors) is truly distinctive.

Finally, carefully recorded crime data stored in a well-designed crime linkage system can also provide the basis for important research studies, only some of which will relate to linkage analysis. For example, data extracted from linkage systems have already been used to conduct interesting studies of serial homicide behavior (e.g., Fritzon & Garbutt, 2001), rape typologies (e.g., McCabe & Wauchope, 2005), criminal profiling (e.g., Kocsis, Cooksey, & Irwin, 2002), and child care providers who commit sexual offenses (e.g., Moulden, Firestone, & Wexler, 2007).

A REVIEW OF THE ASSUMPTIONS UNDERLYING CRIME LINKAGE SYSTEMS AND A RESEARCH AGENDA

Despite the fact that differences can exist in the way crime linkage systems are utilized, even when the same system is used in different jurisdictions, we suggest that these systems

are based on a number of assumptions that override these differences. There are at least four assumptions that are important to consider, given that they will affect the effectiveness of these systems. These assumptions are that (a) data contained in the systems can be coded reliably, (b) data contained in the systems are sufficiently accurate to draw meaningful inferences, (c) violent serial offenders exhibit consistent but distinctive patterns of behavior across their crimes that will enable the linking process, and (d) analysts possess the ability to identify such patterns and link crimes accordingly. In the sections that follow we discuss each of these assumptions and propose future research to examine them. The research agenda we propose can contribute to the effectiveness of linkage systems, both as a research tool and as a tool for use in police investigations.

THE RELIABILITY ASSUMPTION

Evidence for the reliability of data coding. Perhaps the most fundamental assumption underlying the use of crime linkage systems is that the data contained in the systems are reliable. The primary type of reliability of concern here is interrater reliability. A test of interrater reliability in this context would involve determining how often two or more coders (e.g., different investigators) enter the same information into the coding booklet, or system, when applying the same coding categories to the same case. For example, it is assumed that two investigators exposed to the same case material would each record the same occupation for the victim of the crime. In scientific research, a minimum level of 80% agreement is typically required to trust the data on which inferences and conclusions are drawn (e.g., Hartmann, 1977). Arguably, a similarly high level of interrater agreement should be demanded in the law enforcement context, where the inferences being drawn are consequential.

Knowing how reliable data are in this context is critical because the validity of inferences drawn from the data contained in linkage systems depends on a high degree of interrater reliability. Despite the importance of reliability, we are aware of only two studies that have examined this issue. In one study, Martineau and Corey (2008) provided 237 police officers with either a sexual assault or homicide vignette (a two-page summary of the case) and asked them to complete a ViCLAS booklet. The officers were also given the ViCLAS Field Investigator's Guide—a resource that contains explanations of each question found in a ViCLAS booklet—to assist them with their task. Once completed, Martineau and Corey calculated three interrater reliability measures.

In terms of *overall percentage agreement*, Martineau and Corey (2008) reported a rate of 88% agreement for the sexual assault case and 79% agreement for the homicide case—both of which appear acceptable. However, these reliability values are inflated because of the large contribution of *nonoccurrence agreement* between the investigators (i.e., instances where investigators agreed that something did not occur). This is problematic for two related reasons. First, although it is useful for investigators to agree on what did not occur in a case (e.g., that the weapon was *not* a knife, or a bat, or a hammer, or a rock, etc.), it is arguably more important that investigators agree on what did occur (e.g., that the weapon *was* in fact a gun). Thus, the factors contributing to the high level of overall agreement in Martineau and Corey's study are not "equal" in value. Second, the level of overall agreement that is found for a specific variable is dependent on the number of coding options available for that variable. For example, if there are 10 options available for a particular variable under study, such as the type of weapon used, then under a scenario in which only

one option is correct (i.e., the single weapon that *was* used), two investigators are inevitably going to achieve a high level of *overall agreement*. The only two outcomes for the aforementioned example are that officers will agree 8 out of 10 times (80%) or 10 out of 10 times (100%).

In coding situations where there are many opportunities to agree on what *did not happen*, a more sensitive and appropriate measure of interrater reliability is *occurrence agreement* (vs. *nonoccurrence agreement*, or *overall agreement*; see Hartmann, 1977, for an in-depth discussion of these issues). Occurrence agreement is defined as the number of instances where two raters indicate that a particular piece of information was *present* in a vignette (or case file), divided by the total number of instances where at least one of the two observers indicated that a piece of information was *present*, multiplied by 100. When Martineau and Corey (2008) calculated occurrence agreement values, the results were a less impressive 38% agreement for the homicide case and 25% agreement for the sexual assault case. Although certain variable categories were coded in a somewhat reliable fashion (>50% agreement), the majority of categories were not. For the homicide vignette, they found an occurrence agreement of 4% for information about the crime scene, 9% for information about the offense, 13% for information about the offender, 23% for information pertaining to administration questions, 27% for information about the deceased victim, and 32% for information about the victim. Similarly, for the sexual assault vignette, they found an occurrence agreement of 5% for information about the biological sample, 10% for the scene information, 13% for offense information, 13% for offender information, 18% for victim information, and 25% for information pertaining to administration questions. These low percentages demonstrate that officers disagreed with each other about what was *present* in the vignettes more than they agreed.

In a second more recent study, Snook, Luther, House, Bennell, and Taylor (2012) tested 10 police officers to assess the interrater reliability associated with ViCLAS variables. The sample was a relatively homogeneous group of officers from a Canadian police organization. All of the officers investigate ViCLAS-appropriate crimes as part of their job and are in a position to complete ViCLAS booklets. Unlike the study by Martineau and Corey (2008), the officers in this study were provided with a complete case file to code rather than a short vignette. The case file was longer and more detailed than the material used by Martineau and Corey and is more similar to the material that would be coded in naturalistic settings.

For reasons outlined above, Snook et al. (2012) focused on occurrence agreement within their study, and consistent with the results reported by Martineau and Corey (2008), the results indicated low levels of reliability. More specifically, of the 106 variables that were examined in this study, the average level of occurrence agreement was 30.77%. Only 11 (10.38%) of the variables that were coded reached an acceptable level of agreement (i.e., 80%). When the 106 variables were categorized into eight sections, the levels of occurrence agreement ranged from as high as 63% for information pertaining to administrative questions to as low as 2% for information related to weapons. Every category other than the one containing administrative questions was less than 50%, which raises serious questions about whether ViCLAS data can be coded reliably.

If the aforementioned values are representative of the reliability of ViCLAS data, it would be imprudent to draw inferences from that data. It is difficult, however, to determine conclusively whether or not these findings reflect the reliability of ViCLAS data because

both studies are somewhat artificial. Indeed, the values reported above may be an *underestimation* of the true levels of reliability because there was no pressure on the participants to perform in a conscientious manner. Such pressures may be present in naturalistic settings and could increase the effort made by coders and, as a consequence, the reliability of the data. However, equally, the reported values in these two studies may be an *overestimation* of the true values. If investigators cannot achieve reliability when reviewing material under ideal laboratory-type coding conditions, it could be argued that interrater agreement would worsen under naturalistic conditions where, for example, distractions are more common. What is not debatable is the fact that little is currently known about the reliability of the data contained in crime linkage systems such as ViCLAS.

Future research on issues of data reliability. In terms of a research agenda for the future, it is imperative that studies of interrater reliability be conducted for all linkage systems. The studies by Martineau and Corey (2008) and Snook et al. (2012) provide models for conducting such studies. We encourage researchers to place greater emphasis on certain forms of reliability over others when conducting such studies (e.g., occurrence agreement vs. nonoccurrence agreement) and to explore other potential reliability statistics, such as Krippendorff's alpha (Krippendorff, 2004). We also encourage researchers to use research stimuli that match as closely as possible the type of material that would be coded in real-world settings, as Snook et al. did in their recent study, because the amount of investigative material that needs to be processed, and its complexity, could have an impact on the degree of interrater reliability that is achieved. One likely outcome of this type of research will be an enhanced knowledge of the sections contained in coding booklets that might be problematic with respect to interrater reliability. This information could then be used to modify relevant parts of the coding framework or to provide enhanced training on sections identified to be troublesome. Ideally, the impact of these changes would also be evaluated to determine the extent to which they positively affect the degree of interrater reliability that is achieved.

It will also be important to evaluate attempts that are being made by police organizations to increase the reliability of linking data. The RCMP, for example, has created the Field Investigator's Guide for ViCLAS, which attempts to provide investigators with clear definitions of the variables in the ViCLAS coding book, and they have also implemented electronic ViCLAS coding booklets, which may increase the level of interrater reliability associated with ViCLAS data by making the coding task easier and less time-consuming (thereby decreasing coding errors; RCMP, n.d.). Other police organizations have centralized the data coding process (Abraham & O'Dwyer, 2011), which presumably increases the likelihood that the coders have the time, commitment, and expertise to enter the data carefully and correctly. Although these types of changes to the linking process are potentially useful, evaluative research assessing the impact of these innovations on data reliability will be an important undertaking.

THE ACCURACY ASSUMPTION

Evidence for the accuracy of data coding. Another assumption underlying crime linkage systems, at least when applied to certain tasks, is that data entered into the systems accurately reflect what occurred in the criminal event (Martineau & Corey, 2008). In scientific

terms, this is an assumption about the validity of the data: that they represent what they are supposed to represent. This assumption is important because the "quality of the information generated from a database is only as good as the accuracy of the data contained in the database" (Morley & Parker, 2009, p. 599). Although it is technically possible to use reliably coded but invalid data to establish crime linkages, such a system would be operating atheoretically, without any rationale for why the crimes are able to be linked. Without this rationale, it would not be possible to identify the general conditions under which the system will or will not work.[2]

As far as we are aware, there has been no evaluation of the extent to which data stored in linkage systems are valid. Each question included in a coding booklet provides an opportunity for errors to creep into the system, and one ought to be concerned in this setting that the nature of the data coding and entry exercises can potentially result in an increase in errors. For example, the lengthy, repetitive nature of the coding task could result in an unreasonably high number of coding errors (Healy, Kole, Buck-Gengler, & Bourne, 2004), where items that should be coded as being present (or absent) in a crime are incorrectly coded as being absent (or present). This would naturally influence the reliability of the coded data, but it will also negatively affect the accuracy of the data (e.g., making it appear that crimes were committed in a way that they were not). There are factors present in naturalistic police settings that may counteract these problems, such as the level of conscientiousness that may be shown by individuals when coding real cases, but whether or not these factors positively affect the accuracy of linking data is an empirical question that requires testing.

Future research on issues of data accuracy. As is the case with interrater reliability, examinations of data accuracy are urgently required for all linkage systems. Ideally, these examinations will take place in naturalistic settings, using genuine crimes and data coders who are operating under real-world conditions. A useful alternative to field tests, however, is laboratory studies that allow researchers to gauge the degree of data accuracy associated with a particular linkage system. For example, as done routinely in medical settings (e.g., Samuels, Appel, Reddy, & Tilson, 2002), the accuracy of data coding could be tested easily using the details of solved cases. The results of these tests could be used to gauge the extent to which police organizations should trust the data being entered into linkage systems. Such research could also facilitate attempts to maximize data accuracy. For instance, if studies could identify aspects of the data entry process (e.g., variable definitions included in a coding guide) that are related to coding and entry errors, this information could then be used to make improvements. Likewise, such studies could provide quantitative estimates of accuracy for use when evaluating modifications to a linkage system, which are often designed to increase data accuracy.

Most police organizations appear to be aware of issues that may negatively affect data accuracy, and some have even taken steps to minimize their impact. For example, recognizing that some level of human error is inevitable when coding files, numerous organizations have put quality assurance mechanisms in place in an attempt to identify coding errors and correct them before the data are entered into the linkage system (e.g., RCMP, n.d.). We are encouraged by these attempts and think they hold promise. However, evaluative studies are needed to ensure that the mechanisms being put into place are effective. Quality assurance checks will also work only to the extent that they are being utilized reliably and as intended, so the degree to which these checks are being complied with needs to be confirmed.

THE CONSISTENCY AND DISTINCTIVENESS ASSUMPTION

Evidence for consistency and distinctiveness. Albeit implicit, the third assumption made by some developers and users of crime linkage systems is that violent serial offenders will exhibit behaviors across their crimes that are relatively stable and distinctive when compared to behaviors exhibited by other offenders. This assumption is important because, as indicated above, MO behaviors and/or behavioral signatures often seem to be relied on for linking crimes (e.g., Gault, 2010; Kangas, 2001; Keppel, 2000; Keppel et al., 2005; Martineau & Corey, 2008; Ressler et al., 1988). For example, as Martineau and Corey (2008) state in their study of ViCLAS, "[W]hen an analyst identifies a number of cases that share significant behavioral similarities . . . the analyst will link the cases to form a potential series" (p. 52).

The assumption that offenders will exhibit a distinctive MO across their crimes appears to originate from the view that behavior is determined primarily by internal traits, or dispositions to behave in a particular way (Cervone & Shoda, 1999). If personal traits are the primary determinant of behavior, it would be reasonable to expect people to exhibit distinctive patterns of behavior in a stable fashion across situations. However, research has demonstrated that situational factors also play a key role in determining how people behave, which explains the low levels of behavioral consistency that are often found (Mischel, 1968). When stable patterns of behavioral distinctiveness are found in the noncriminal domain, they are most often found across situations that are viewed as psychologically similar by the individual being observed (e.g., Shoda, Mischel, & Wright, 1994), or for behaviors that are largely under the control of the individual, rather than a product of the situation (e.g., Funder & Colvin, 1991).

Although only limited research exists for the types of violent crimes included in most crime linkage systems, a criminal's MO, much like his or her noncriminal behavior, appears to be determined by both personal preferences to behave in a particular way and a range of situational factors (Woodhams, Hollin, & Bull, 2008).[3] For example, although a serial rapist may have a preferred behavioral style when committing his or her crimes (e.g., a disposition favoring pseudo-intimate interactions; Canter, Bennell, Alison, & Reddy, 2003), the behavior of the offender may change (e.g., become more hostile) if he or she experiences a high level of victim resistance in a particular crime. It should come as no surprise then that although very high levels of consistency are sometimes reported in the literature (e.g., Melnyk, Bennell, Gauthier, & Gauthier, 2011), the majority of research examining the MOs of violent offenders has found low to moderate levels of behavioral consistency (e.g., Bateman & Salfati, 2007; Bennell, Jones, & Melnyk, 2009; Grubin et al., 2001; Santtila et al., 2008; Santtila, Junkkila, & Sandnabba, 2005; Sjöstedt, Långström, Sturidsson, & Grann, 2008). For example, Sjöstedt et al. (2008) used kappa statistics to assess the temporal stability of MOs exhibited by 75 Swedish sex offenders who recidivated. When comparing each offender's prior sex offense to his first reoffense, low kappa scores were found for the majority of MO features that were examined (e.g., ranging from $\kappa = .22$ to $\kappa = .34$ for variables such as noncontact offense, physical contact, penetration, death threat, and victim injury). The only MO feature that was relatively stable was victim choice, although stability varied substantially as a function of victim type (ranging from $\kappa = .79$ for male victim to $\kappa = .49$ for stranger victim). Given these sorts of results, linkage analysts must be extremely cautious when relying on MO indicators to link violent crimes until research emerges that can inform their selection of useful linking behaviors.

The assumption that signature behaviors may be useful for linkage analysis seems to be based on the belief that signatures instantiate offenders' "scripts" that are typically well rehearsed, deeply engrained, and rooted in personal fantasies (Canter & Heritage, 1990; Davies, 1992; Hazelwood & Warren, 1990). In contrast to MO, we are not aware of any empirical, published research that has examined the potential value of behavioral signatures for linking crimes (at least not using commonly accepted definitions of behavioral signatures; see Bateman & Salfati, 2007). Although case studies have been presented to support the idea that signature behaviors can be identified and used to link serious violent crimes (e.g., Douglas & Munn, 1992; Keppel, 1995, 2000; Keppel & Birnes, 1997; Keppel et al., 2005), such anecdotes may not be generalizable.

Of course it might be possible to link crimes without relying on MO or behavioral signatures. Indeed, other data that are stored in linkage systems, such as offender descriptions or victim information, could potentially be used for this purpose. However, to the extent that behavioral information is relied on, caution is warranted when considering the links that are established. The potential dangers of using behavioral information for linking purposes should be made clear to the individuals who carry out linkage analysis and to the investigators provided with the results of such analysis. Based on our knowledge of Canadian training programs (e.g., for ViCLAS analysts), this message is being delivered to the people involved in conducting the analysis. However, it is unclear whether this message is influencing the way in which analysts operate and the subsequent advice being passed on to investigators.

Future research on issues of consistency and distinctiveness. As we have just argued, the task of linking crimes to a common offender sometimes depends on there being evidence that offenders display a relatively high level of behavioral consistency and distinctiveness across the crimes they commit. Although some research has examined this issue in relation to MO, we are aware of no empirical research examining behavioral signatures, as traditionally defined. Indeed, we still do not know whether linkage analysts can identify behavioral signatures across various types of crimes, if they are in fact exhibited, or the extent to which these signatures are useful for establishing crime linkages. Research in this area should build on existing research (see Woodhams, Hollin, & Bull, 2007, for a review) by striving to identify the conditions under which consistency and distinctiveness will be found (for both MO and signatures). This research should be conducted using a range of crime types, especially those crimes that are entered into crime linkage systems (e.g., violent interpersonal crimes). Clearly, more research also has to be conducted to determine how the analysis of crime scene behaviors (the primary focus in most linking research) can be integrated into the analysis of other potential linking factors, such as physical evidence, offender characteristics, or victim descriptions.

Research of the type described in the previous paragraph could inform the construction of more streamlined coding books, which might persuade more investigators to complete them (a common problem with many linkage systems; RCMP, n.d.). This research will also inform linkage analysts about what behaviors to focus on when linking crimes, and in what order. Finally, this research could result in algorithms that accomplish much of the linking work for the analyst, in a similar way to what is happening in the risk assessment field (e.g., Quinsey, Harris, Rice, & Cormier, 2006). It is interesting that attempting to derive such algorithms provides a complementary test of the capacity of analysts to link crimes, since the absence of any successful algorithm suggests that no linear combination of evidence

could be used for this purpose. This finding would demand that some penetrating questions be asked of how exactly analysts are undertaking and achieving the linking task successfully, assuming that they are.

THE ABILITY ASSUMPTION

Evidence for linking ability. The last assumption is that people who have received specialized training to link crimes possess the ability to identify serial crimes contained in linkage systems. We are not aware of any research that has examined the degree to which trained linkage analysts can make linking decisions accurately. Nor do we know of any research that has examined performance in the types of linking tasks that analysts actually face in naturalistic settings (e.g., the factors on which linking decisions are based in existing studies are almost always restricted to crime scene behaviors, which represent only a subset of the variables available to analysts in the real world, and the samples of crimes that are presented to participants in these studies often bear little resemblance to the complex samples of linked, unlinked, and one-off crimes that linkage analysts have to contend with). The only research available that provides a sense for how effective people are at identifying linked crimes has been conducted using law enforcement personnel (and members of the public) who have not received formal training in linkage analysis. As indicated above, this research also tends to use linking tasks that are relatively low in ecological validity.

The first of these studies was conducted by Canter and his colleagues (1991). Canter et al. provided 32 detectives in the United Kingdom the crime scene descriptions of 12 sexual attacks committed by four known offenders (three crimes per offender). Their task was to read the descriptions, identify features of the crimes that are useful for linking purposes, and decide which of the crimes were linked. Out of a possible 12 correct links, the modal result (10 detectives) was 3 correct links. One detective identified no correct links, and the highest number of correct links, which were identified by three detectives, was 8. The results for all the other detectives in the study fell between these two extremes.

A similar study by Santtila, Korpela, and Häkkänen (2004) examined the ability of four distinct groups of individuals to link vehicle offenses accurately. They presented experienced vehicle offense investigators, experienced general investigators (i.e., investigators with no specialized training in vehicle crime investigation), novice general investigators, and naive participants with offense information relating to 30 offenses committed by 10 known offenders (3 offenses each). The participants were asked to review the offense information and determine which of the offenses were linked. They found that investigators were significantly more accurate than naive participants, but there were no differences in accuracy between different types of investigators (each group identified about half of the possible links correctly).

Most recently, Bennell, Bloomfield, Snook, Taylor, and Barnes (2010) examined how university students, police professionals, and a logistic regression model performed on a linking task. Information on 38 pairs of burglaries, some of which represented linked crimes, was provided to each participant. Half of the participants in each group were provided with training that informed them that the likelihood of two offenses being committed by the same offender increases as the distance between the offenses decreases (see Bennell & Canter, 2002). Participants were asked to decide for each offense pair whether or not the same offender committed the crimes. They found that students outperformed police professionals, that providing information about appropriate linking cues can increase accuracy

significantly, but that statistical models tend to outperform human judgment. The major problem for the participants, even in the trained condition, was an overreliance on ineffective linking cues, which seemed to result from inaccurate beliefs about what MO features in burglary are consistent and/or distinctive.

Accepting that there are some issues with their external validity, these studies do not provide strong evidence that various types of police professionals, including experienced investigators, can link serial crimes accurately. If the studies just reviewed had shown positive results, it would be reasonable to assume that trained linkage analysts would do as well as the tested participants, if not better, since they have been trained in the linkage task. Unfortunately, the reality is that we simply do not know at present whether trained linkage analysts possess the ability to link serial crimes, either in laboratory-based studies or in naturalistic settings where systems such as ViCLAS are used.

Another concern is that the poor performance of participants in linking studies raises the possibility that the fallibility of human decision making, which is often found in other decision-making domains (Jacob, Gaultney, & Salvendy, 1986; Kahneman, Slovic, & Tversky, 1982; Kleinmuntz, 1990), may be a problem in the investigative domain too. If this is true, trained linkage analysts may fare little better than the participants in these studies. What we know about how actual linking decisions are made in some operational settings by trained analysts does little to ease our concerns. For example, we have been informed that many trained linkage analysts rely on an experience-based, subjective, idiographic approach for selecting linking cues rather than a data-driven, objective, nomothetic approach (also see RCMP, n.d.). This does not accord well with the published decision-making literature, which has historically highlighted the superiority of the latter approach over the former (e.g., Grove & Meehl, 1996). Indeed, that literature suggests that (unaided) linkage analysts may be ill equipped to perform well on the linking task, thus making it difficult for them to achieve accuracy rates that could be achieved using more mechanical (standardized) decision-making procedures.

Future research on issues of linking ability. No matter what emerges from any of the other lines of research described above, the ultimate issue from an operational policing perspective will be whether or not linkage analysts can establish links between crimes accurately. Although it would be ideal to conduct field studies to examine linkage decision accuracy, a more fruitful alternative, at least initially, may be to conduct laboratory tests that use specific linkage systems. The studies described above (e.g., Bennell et al., 2010; Canter et al., 1991; Santtila et al., 2004) provide one potential approach for conducting such studies, and the degree of external validity associated with these studies could be improved with the assistance of police organizations. The findings from such studies could increase the degree of linking accuracy that can be achieved.

Some obvious questions for future research in this area might be the following: What linking strategies are currently used by linkage analysts? What level of accuracy can be achieved when using these strategies? How does this level of accuracy compare to the level of accuracy achieved using other (e.g., actuarial) methods? Can the accuracy of linking decisions be improved with the introduction of additional decision support tools or empirically informed training? What factors (e.g., experience, training, motivation, pressure) influence one's ability to make accurate linking decisions?

OTHER IMPORTANT RESEARCH

Beyond research examining the aforementioned assumptions, another major research endeavor should be to establish the success rate (e.g., number of actual links established) associated with the various linkage systems being used by police organizations. Although we are aware of anecdotes of successful police investigations that have drawn on linkage systems as part of their investigations, these anecdotes by themselves do not constitute strong evidence in support of linkage systems. We are not aware of any data from wide-scale studies that provide an indication of how successful existing linkage systems are in assisting investigations or the role that linkage systems played in investigative successes. This is despite the fact that some of these systems are meant to be updated when potential links are confirmed or rejected by investigators.

Future research that establishes the effectiveness of linkage systems will be useful for at least two reasons. First, it will provide developers of these systems, the analysts who use them, and the investigators who depend on the results with evidence that the systems are in fact effective and worth the cost (in terms of time, effort, and money). Second, this research will provide a baseline with which to compare future adjustments to the coding framework, the system itself, and the analysts using the system (e.g., with respect to training). We recommend that this research go beyond an evaluation of *potential* links, for which some information is already available. The real need is to establish the number of *actual* links achieved by the system under investigation. Obtaining an accurate measure of system success requires that investigators furnished with a potential link provide feedback on whether or not the crimes were actually linked. Getting such feedback from time-constrained investigators will be challenging, but actual links is clearly the best measure of a system's success.

A FINAL WORD

Our intention with this article was not to minimize the efforts of law enforcement personnel who have dedicated much time and energy to developing crime linkage systems, nor was it our intention to criticize those analysts who work tirelessly to identify serial offenders. It was our intention to examine critically the assumptions underlying crime linkage systems, systems that are widely used around the world without being subjected to empirical scrutiny. Police agencies will have to decide for themselves how much weight to put on the issues we raised in this article and what conclusions to draw regarding the potential value of linkage systems. At the very least, we believe that these systems ought to be evaluated as a matter of urgency because there exists a real risk that current linking efforts are not achieving optimal results. We hope the research agenda described in this article, if adopted by researchers, will go some way toward addressing the concerns raised in this critical review and allow crime linkage systems (and the analysts who operate them) to reach their full potential.

NOTES

1. For example, the 38-page Violent Crime Linkage Analysis System booklet currently consists of 156 open-ended and closed-ended questions that relate to a number of categories, such as victim, offender, and offense information. There is also an e-version of this booklet that is now available.

2. Of course, the value of any research that draws on data contained within linkage systems ultimately hinges on the assumption of data accuracy as well. Exactly how accurate such data need to be to be useful for research or linking purposes is a difficult question to answer, but the importance of data accuracy, in general terms, is surely something that everyone can agree on.

3. Issues of consistency and distinctiveness are often examined in relation to property crimes, such as arson, burglary, and vehicle theft, which are rarely included in linkage systems (e.g., Bennell & Canter, 2002; Bennell & Jones, 2005; Ewart, Oatley, & Burn, 2005; Goodwill & Alison, 2006; Markson, Woodhams, & Bond, 2010; Santtila, Fritzon, & Tamelander, 2004; Tonkin, Grant, & Bond, 2008; Woodhams & Toye, 2007).

REFERENCES

Abraham, L., & O'Dwyer, J. (2011, November). *Crime series linkage*. Paper presented at the International Crime and Intelligence Analysis Conference, Manchester, UK.

Bateman, A. L., & Salfati, C. G. (2007). An examination of behavioral consistency using individual behaviors or groups of behaviors in serial homicide. *Behavioral Sciences & the Law*, *25*, 527-544. doi:10.1002/bsl.742

Bennell, C., Bloomfield, S., Snook, B., Taylor, P. J., & 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 and Law*, *16*, 507-524. doi:10.1080/10683160902971030

Bennell, C., & Canter, D. (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., & 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

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. doi:10.1016/j.ijlp.2010.03.009

Canter, D. V., Bennell, C., Alison, L. J., & Reddy, S. (2003). Differentiating sex offences: A behaviourally based thematic classification of stranger rapes. *Behavioral Sciences & the Law*, *21*, 157-174. doi:10.1002/bsl.526

Canter, D. V., & Heritage, R. (1990). A multivariate model of sexual offence behaviour: Developments in "offender profiling." *Journal of Forensic Psychiatry*, *1*, 185-212. doi:10.1080/09585189008408469

Canter, D. V., Heritage, R., Wilson, M., Davies, A., Kirby, S., Holden, R., McGinley, J., Hughes, H., Larkin, P., Martin, L., Tsang, E., Vaughan, G., & Donald, I. (1991). *A facet approach to offender profiling*. London, UK: Home Office.

Cervone, D., & Shoda, Y. (1999). Beyond traits in the study of personality coherence. *Current Directions in Psychological Science*, *8*(1), 27-32. doi:10.1111/1467-8721.00007

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

Cooper, G. (2007). The Violent Criminal Apprehension Program. *Forensic Examiner*, *16*, 72-73. Retrieved from http://findarticles.com/p/articles/mi_go1613/is_3_16/ai_n29371118/?tag=content;col1

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

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

Egger, S. (1984). A working definition of serial murder and the reduction of linkage blindness. *Journal of Police Science and Administration*, *12*, 348-357.

Ewart, B. W., Oatley, G. C., & Burn, K. (2005). Matching crimes using burglars' modus operandi: A test of three models. *International Journal of Police Science and Management*, *7*, 160-174.

Federal Bureau of Investigation. (n.d.). *Investigations and operations support*. Retrieved from http://www.fbi.gov/about-us/cirg/investigations-and-operations-support

Fritzon, K., & Garbutt, R. (2001). A fatal interaction: The role of the victim and function of aggression in intrafamilial homicide. *Psychology, Crime and Law*, *7*, 309-331. doi:10.1080/10683160108401800

Funder, D., & Colvin, C. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology*, *60*, 773-794. doi:10.1037/0022-3514.60.5.773

Gault, K. (2010). To catch a killer: Profiling based on experience, not a crystal ball. *RCMP Gazette*, *72*(2), 10.

Goodwill, A., & Alison, L. (2006). The development of a filter model for prioritizing suspects in burglary offences. *Psychology, Crime and Law*, *12*, 395-416. doi:10.1080/10683160500056945

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293-323. doi:10.1037/1076-8971.2.2.293

Grubin, D., Kelly, P., & Brunsdon, C. (2001). *Linking serious sexual assaults through behaviour*. London, UK: Home Office.

Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, *10*, 103-116. doi:10.1901/jaba.1977.10-103

Hazelwood, R. R., & Warren, J. (1990). The criminal behavior of the serial rapist. *FBI Law Enforcement Bulletin*, *59*(2), 1-17.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied*, *10*, 188-199. doi:10.1037/1076-898X.10.3.188

Jacob, V. S., Gaultney, L. D., & Salvendy, G. (1986). Strategies and biases in human decision making and their implications for expert systems. *Behaviour and Information Technology*, *5*, 119-140. doi:10.1080/01449298608914505

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.

Kangas, L. (2001). *Artificial neural network system for the classification of offenders in murder and rape cases*. Washington, DC: National Institute of Justice.

Keppel, R. D. (1995). Signature murders: A report of several related cases. *Journal of Forensic Sciences*, *40*, 670-674.

Keppel, R. D. (2000). Signature murders: A report of the 1984 Cranbrook, British Columbia Cases. *Journal of Forensic Sciences*, *45*, 500-503.

Keppel, R. D., & Birnes, W. (1997). *Signature killers*. New York, NY: Pocket Books.

Keppel, R. D., Weis, J. G., Brown, K. M., & Welch, K. (2005). The Jack the Ripper murders: A modus operandi and signature analysis of the 1888–1891 Whitechapel murders. *Journal of Investigative Psychology and Offender Profiling*, *2*, 1-21. doi:10.1002/jip.22

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, *107*, 296-310.

Kocsis, R., Cooksey, R., & Irwin, H. (2002). Psychological profiling of sexual murders: An empirical model. *International Journal of Offender Therapy and Comparative Criminology*, *46*, 532-554. doi:10.1177/030662402236739

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Labuschagne, G. (2006). The use of a linkage analysis as evidence in the conviction of the Newcastle serial murderer, South Africa. *Journal of Investigative Psychology and Offender Profiling*, *3*, 183-191. doi:10.1002/jip.51

Markson, L., Woodhams, J., & Bond, J. (2010). Linking serial residential burglary: Comparing the utility of modus operandi behaviours, geographical proximity and temporal proximity. *Journal of Investigative Psychology and Offender Profiling*, *7*, 91-107. doi:10.1002/jip.120

Martineau, M., & Corey, S. (2008). Investigating the reliability of the Violent Crime Linkage Analysis System (ViCLAS) crime report. *Journal of Police and Criminal Psychology*, *23*, 51-60. doi:10.1007/s11896-008-9028-5

McCabe, M., & Wauchope, M. (2005). Behavioral characteristics of men accused of rape: Evidence for different types of rapists. *Archives of Sexual Behavior*, *34*, 241-253. doi:10.1007/s10508-005-1801-2

Melnyk, T., Bennell, C., Gauthier, D., & Gauthier, D. (2011). Another look at across-crime similarity coefficients: An attempt to replicate Woodhams, Grant, and Price (2007). *Psychology*, *Crime and Law*, *17*, 359-380. doi:10.1080/10683160903273188

Meyer, C. B. (2007). Criminal profiling as expert evidence. In R. N. Kocsis (Ed.), *Criminal profiling: International theory*, *research*, *and practice* (pp. 207-247). Totowa, NJ: Humana Press.

Mischel, W. (1968). *Personality and assessment.* New York, NY: John Wiley.

Morley, D., & Parker, C. S. (2009). Databases and database management systems. In D. Morley & C. S. Parker (Eds.), *Understanding computers: Today and tomorrow* (pp. 588-627). Boston, MA: Cengage Learning.

Moulden, H., Firestone, P., & Wexler, A. (2007). Child care providers who commit sexual offences: A description of offender, offence, and victim characteristics. *International Journal of Offender Therapy and Comparative Criminology*, *51*, 384-406. doi:10.1177/0306624X06298465

Ormerod, D., & Sturman, J. (2005). Working with the courts: Advice for expert witnesses. In L. J. Alison (Ed.), *The forensic psychologist's casebook: Psychological profiling and criminal investigation* (pp. 170-193). Devon, UK: Willan.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Ressler, R. K., Burgess, A. W., & Douglas, J. E. (1988). *Sexual homicide: Patterns and motives*. Lexington, MA: Lexington.

Royal Canadian Mounted Police. (n.d.). *RCMP Violent Crime Linkage System (ViCLAS)*. Retrieved from http://www.rcmp-grc.gc.ca/viclas-salvac/index-eng.htm

Samuels, R. C., Appel, L., Reddy, S. I., & Tilson, R. S. (2002). Improving accuracy in a computerized immunization registry. *Ambulatory Pediatrics*, *2*, 187-192. doi:10.1367/1539-4409(2002)002<0187:IAIACI>2.0.CO;2

Santtila, P., Fritzon, K., & Tamelander, A. L. (2004). Linking arson incidents on the basis of crime scene behaviour. *Journal of Police and Criminal Psychology*, *19*(1), 1-16. doi:10.1007/BF02802570

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 linking of car crime series. *Psychology*, *Crime and Law*, *10*, 97-112. doi:10.1080/1068316021000030559

Santtila, P., Pakkanen, T., Zappalá, A., Bosco, D., Valkama, M., & Mokros, A. (2008). Behavioural crime linking in serial homicide. *Psychology*, *Crime and Law*, *14*, 245-265. doi:10.1080/10683160701739679

Shoda, Y., Mischel, W., & Wright, J. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, *67*, 674-687. doi:10.1037/0022-3514.67.4.674

Sjöstedt, G., Långström, N., Sturidsson, K., & Grann, M. (2008). Stability of modus operandi in sexual offending. *Criminal Justice and Behavior*, *31*, 609-623. doi:10.1177/0093854804267094

Snook, B., Luther, K., House, J., Bennell, C., & Taylor, P. J. (2012). The violent crime linkage analysis system: A test of interrater reliability. *Criminal Justice and Behavior*, *39*, 607-619.

State v. Code, 627 So.2d 1373 (1994).

State v. Fortin, 178 N.J. 540; 843 A.2d 974 (2004).

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

State v. Prince, 9 CAL.APP.4th 1176, 10 CAL.RPTR.2D 855 (1992).

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

Witzig, E. W. (2003). New ViCAP: More user-friendly and used by more agencies. *FBI Law Enforcement Bulletin*, *72*, 1-7.

Woodhams, J., Hollin, C., & 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., & 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

Craig Bennell is an associate professor of psychology at Carleton University and director of Carleton's Police Research Lab. Much of his research examines the reliability, validity, and usefulness of psychologically based investigative techniques, including methods that are used to link serial crimes. His other stream of research examines factors that influence police decision making in use-of-force encounters.

Brent Snook is an associate professor of psychology and director of the Bounded Rationality and the Law Laboratory at Memorial University of Newfoundland. His research primarily involves ways of improving the criminal justice system, which includes the study of decision making in legal settings, interviewing and interrogations, and pseudoscience in the criminal justice system.

Sarah MacDonald is a PhD student at Memorial University of Newfoundland, Canada. She obtained her MSc at Memorial University of Newfoundland in 2011 and her BSc (honours) at Dalhousie University in 2009. Her general research interests include investigative interviewing practices and deception detection.

John C. House is a superintendent of the Royal Newfoundland Constabulary (RNC) in Canada. He holds an MSc in investigative psychology from the University of Surrey and presently directs the Criminal Investigation Division of the RNC. He has an interest in the critical evaluation of approaches and methods used by police organizations, with a view that best practices are implemented.

Paul J. Taylor is senior lecturer (associate professor) in the Department of Psychology at Lancaster University, UK. He directs the university's Investigative Expertise Unit, a multidisciplinary research group that brings together expertise in behavioral science and technology to address areas such as protective security, vulnerable witness interviewing, and the investigation of terrorism.