

The Violent Crime Linkage Analysis System: A Test of Inter-Rater Reliability

Brent Snook and Kirk Luther

Memorial University

John C. House

Royal Newfoundland Constabulary

Craig Bennell

Carleton University

Paul J. Taylor

Lancaster University

Author Note

Brent Snook and Kirk Luther, Department of Psychology, Memorial University of Newfoundland, St. John's, NL, Canada; John C. House, Royal Newfoundland Constabulary, St. John's, NL, Canada; Craig Bennell; Department of Psychology, Carleton University, Ottawa, ON, Canada; Paul J. Taylor, Department of Psychology, Lancaster University, Lancaster, U.K.

Correspondence concerning this article should be addressed to Brent Snook, Department of Psychology, Science Building, Memorial University of Newfoundland, St. John's, NL, Canada, A1B 3X9. E-mail: bsnook@mun.ca or brent.snook@gmail.com.

Abstract

The inter-rater reliability of an internationally renowned crime linkage system – the Violent Crime Linkage Analysis System (ViCLAS) – was tested. Police officers ($N = 10$) were presented with a case file and asked to complete a ViCLAS booklet. The level of occurrence agreement between each officer was calculated. Results showed a 30.77% level of agreement across the 106 variables examined. Agreement ranged from 2.36% for weapon variables to 62.87% for administration variables. Only 11 (10.38%) of the variables reached an acceptable level of agreement. Concerns pertaining to the validity of inferences produced using ViCLAS data are discussed, along with potential explanations for the findings, limitations of the study, and future research directions.

Keywords: Violent Crime Linkage Analysis System, Comparative Case Analysis, Serial Crimes, Behavioural Linking, Policing, Inter-Rater Reliability

The Violent Crime Linkage Analysis System: A Test of Inter-Rater Reliability

The task of determining whether the same offender has committed multiple crimes is important to both the effectiveness and efficiency of police investigations (Grubin, Kelly, & Brunsdon, 2001). The ability to perform this linking task accurately often depends on the availability of trace or biological evidence (e.g., fingerprints, DNA). In the absence of such evidence, other evidence such as behavioral information must be relied upon (e.g., verbal exchanges between offender and victim; Woodhams, Hollin, & Bull, 2007). To obtain the behavioral information required to link crimes, law enforcement agencies in North America and much of Europe use computerized linkage systems that are populated with information that officers extract from their case files (Bennell, Snook, MacDonald, House, & Taylor, 2011).

Currently, the most widely used linkage system is the Violent Crime Linkage Analysis System (ViCLAS). A system developed by the Royal Canadian Mounted Police (RCMP; Collins, Johnson, Choy, Davidson, & MacKay, 1998), ViCLAS is used within Canada, and is reportedly being used in Australia, Austria, Belgium, England, France, Germany, the Netherlands, New Zealand, Switzerland, and in two US states (i.e., Indiana and Tennessee; RCMP, n.d.). Although ViCLAS has existed for nearly 20 years, there is no published account of its effectiveness being evaluated systematically. In the current paper, we test one of the key assumptions underlying ViCLAS, namely, that data included in the system may be coded reliably (see Bennell et al., 2011). Before examining the issue of reliability, we will outline the origins of ViCLAS and describe some of the ways that the system is (and potentially may be) used.

Origins and Use of ViCLAS

The origins of computerized crime linkage systems, including ViCLAS, can be traced back to the FBI's Violent Criminal Apprehension Program (VICAP; Collins et al., 1998); a system developed to avoid "linkage blindness," which is the lack of communication between law enforcement agencies across jurisdictional boundaries that can inhibit the apprehension of serial offenders (Clark, 2002; Egger, 1984). The RCMP subsequently developed their own crime linkage system known as the Major Crimes File (MCF; RCMP, n.d.), which was later replaced by ViCLAS (Clark, 2002). It has been argued that use of ViCLAS has become widespread since its beginnings in the 1990s, and that it has gained a reputation as the best crime linkage system available (Bijleveld, & Smit, 2006; Collins et al., 1998).

Although there are variations in how ViCLAS is used across jurisdictions within and between countries, the procedure generally involves five steps (RCMP, n.d.). First, an officer enters data collected from their investigation of a ViCLAS-appropriate crime (e.g., abduction, sexual assault, homicide) into a 38-page booklet or e-booklet. It is important to note that training is not required before officers are allowed to complete ViCLAS booklets. Second, completed booklets are forwarded to a quality control center to be reviewed (Friesen, 2004). Third, data are entered into the database of previously solved and unsolved cases. Fourth, the data are compared against the ViCLAS database by analysts who are trained to search for potential links. Fifth, once the search for linked crimes is complete, relevant investigators are informed about any potential links that have been identified. These investigators then attempt to confirm or eliminate the potential link through further investigation.

In addition to assisting with investigations, ViCLAS has the potential to be used in at least two other ways. One additional use of ViCLAS relates to data storage and management, often for the purpose of conducting research (e.g., on serial homicide behavior, Bijleveld & Smit,

2006; rape typologies, McCabe & Wauchope, 2005; criminal profiling, Kocsis, Cooksey, & Irwin, 2002; and child care providers who commit sexual offences, Moulden, Firestone, & Wexler, 2007). A second use stems from ViCLAS's potential to be used as the basis for similar fact evidence in court cases where questions arise about whether or not a defendant is responsible for multiple crimes (Ormerod & Sturman, 2005). On several occasions, law enforcement officers have provided expert testimony about the likelihood of a series of crimes having been committed by the same offender (e.g., Labuschagne, 2006; *State v. Code*, 1994; *State v. Pennell*, 1989; *State v. Prince*, 1992). Crime linkage systems, such as ViCLAS, may be used to support such testimony (*State v. Fortin*, 2004).

A Preliminary Test of the Reliability of ViCLAS

In order for the data contained in ViCLAS to be of value for the aforementioned purposes, the assumptions upon which it is based must be valid. Perhaps one of the most fundamental assumptions underlying ViCLAS (and all other all linkage systems) is that the database contains reliable data (for a review of the central assumptions underlying ViCLAS see Bennell et al., 2011). The primary type of reliability of concern is inter-rater reliability. A test of inter-rater-reliability involves determining the extent to which two (or more) different investigators enter the same information about a case into a ViCLAS coding booklet (e.g., both investigators agree that the victim was a nurse). In scientific research, a minimum level of 80% agreement is typically deemed acceptable before the inferences and conclusions drawn from coded data are trusted (Hartmann, 1977). It seems logical that a similarly high level of agreement should be demanded from crime linkage systems because of the consequential nature of the inferences drawn from the data held in these systems.

To date, the only test of the reliability of ViCLAS is Martineau and Corey's (2008) study.¹ They provided 237 police officers with a vignette of either a sexual assault or homicide case, and asked them to complete a ViCLAS booklet (see Martineau & Corey's Appendix A and B for the information provided to participants). The participants were also given the ViCLAS field investigator's guide – a resource that explains each of the questions in the booklet – to assist them with the task. Once completed, Martineau and Corey calculated three measures of inter-rater reliability, namely, overall percentage agreement, percentage occurrence agreement, and percentage non-occurrence agreement. In terms of overall percentage agreement, they reported a rate of 88% for the sexual assault case and 79% for the homicide case. Based largely on these results, they concluded that the data contained within ViCLAS was likely to be reliable.

However, in this case, overall percentage agreement inflates the actual agreement between individuals because it is influenced largely by the high levels of non-occurrence agreement between investigators (i.e., instances where both investigators did not record the occurrence of a variable). Consider the example where investigators have five options available to them in the ViCLAS booklet when coding blunt force trauma. If we assume for a moment that the investigators disagree on the type of trauma that was present in a case, they would agree on three of the options (the non-occurrences) and disagree on two of the options (the occurrences). Thus, they would exhibit a 60% agreement rate despite not agreeing on the behavior being coded (i.e., actual type of trauma). Although it is useful that investigators agreed on what did not occur in a particular case (e.g., that the blunt force was not minimal), it is more important from an operational standpoint that investigators agree on what actually happened (e.g., that the blunt force was extreme). In coding situations where multiple options cannot logically be selected simultaneously, such that there is inevitable agreement on what did not occur, the most

appropriate measure of agreement is percentage occurrence agreement. Percentage occurrence agreement captures the number of instances that two coders indicated that a particular piece of information was present in a case file (Hartmann, 1977).

When percentage occurrence agreement values were calculated, Martineau and Corey's (2008) reliability scores were less impressive. Specifically, they reported an overall percentage occurrence agreement of 38% for the homicide case and 25% for the sexual assault case. For the homicide case, occurrence agreement was approximately 4% for crime scene information, 9% for offence information, 13% for offender information, 23% for information pertaining to administration questions, 27% for deceased victim information (a subsection of the offence section), and 32% for information about the victim. Two of the sections (vehicle and weapon) exceeded 80% agreement, which Hartmann (1977) recommended as a desired benchmark. Similarly, for the sexual assault case, occurrence agreement was approximately 5% for information about the biological sample, 10% for the scene information, 13% for both offence and offender information, 18% for victim information, and 25% for the administration questions. The weapon section exceeded the 80% level of agreement.

The low percentage occurrence agreement values reported by Martineau and Corey (2008) demonstrate that officers almost always disagreed with each other about what occurred in certain aspects of the case. Although only a single study, their findings suggest the opposite of what they concluded. Rather than supporting the reliability of ViCLAS data, the results actually indicate that many sections of ViCLAS data are likely to be unreliable and raise concerns about the validity of the inferences derived from using this data. That being said, there are several limitations associated with this study that make it inappropriate to draw any strong conclusions about the operational value of ViCLAS.

One of the most obvious limitations is that Martineau and Corey (2008) did not use actual case files in their study, preferring instead to rely on short vignettes. Thus, their results were based on less detailed information than would normally be the available in more naturalistic settings where officers have access to larger amounts of investigative material. A second limitation is that their participants consisted of a relatively heterogeneous group of police officers who may have differed from each other in important ways (e.g., whether they serve in a role that requires the completion of ViCLAS booklets).

The Current Study

The goal of the current study is to conduct an independent assessment of the inter-rater reliability of ViCLAS data in a manner similar to Martineau and Corey (2008), while also addressing some of the methodological limitations inherent in that previous study. We examine the level of percentage occurrence agreement associated with a wide range of ViCLAS variables using actual case file materials that are longer and more detailed than the previously used vignettes. We also sample a more homogeneous group of police officers by testing officers who investigate ViCLAS-appropriate crimes, and are in a position to complete ViCLAS booklets. These changes to Martineau and Corey's design will allow us to determine the degree of inter-rater reliability associated with ViCLAS data in a more ecologically valid fashion.

Method

Participants

Participants were 10 police officers (7 men and 3 women) working as crime investigators in a Canadian police organization. They completed the study voluntarily and did not receive any compensation. One participant reported being in the age range of 26-30 years, one in the 31-35 year age range, three in the 36-40 year age range, four in the 41-45 year age range, and one

reported being older than 45 years of age. Four participants indicated that they had between 5 to 10 years of policing experience and six indicated that they had more than 10 years of experience.

In response to the open-ended question about their experiences with ViCLAS, nine participants reported having completed ViCLAS booklets. Three participants reported having spent less than one year of service completing booklets as part of their job, and the remaining seven indicated spending between two to eight years of service completing booklets as part of their job. Seven participants indicated that they did not complete any booklets in the past year, two reported completing between one to two booklets per week in the past year, and one reported completing four booklets in the past year. Additionally, two participants reported spending less than one hour per week completing booklets in the last year, one reported dedicating between two to three hours per week on the task, and seven stated that they did not dedicate any hours to the task in the past year. Six participants reported not having any previous training on how to complete ViCLAS booklets and the other four indicated receiving an introduction to the task during their cadet training.

Participants were also asked six questions about the field investigator's guide. Only one participant reported that they were aware that the guide existed, one participant reported using the guide previously, none of the participants indicated using the guide routinely, and only one participant reported receiving training on how to use the guide. When asked to rate their familiarity with the field investigator's guide on a 10-point scale, where *1 = not at all familiar* and *10 = extremely familiar*, the average familiarity score was 1.11 (*SD* = 0.33). None of the participants indicated being aware of any successful links based on ViCLAS booklets they completed.

Materials

A genuine case file from a solved criminal investigation was used as the stimulus for the current study. The file contained an anonymous 29-page victim interview transcript and a three page case report from a case involving forcible confinement and sexual assault. The case report contained details of the accused (i.e., name, gender, date of birth, height, weight, previous offences) and a summary of the criminal events. The information included in the transcript and case report provided information necessary for completing Version 4.0 of the ViCLAS booklet (e.g., descriptors and behaviors of the victim and offender, aspects of the crime scene). On a 5-point realism scale, where *1 = not realistic* and *5 = very realistic*, participants average rating of the realism of the experimental material was 4.00 ($SD = 1.25$).

The 38-page ViCLAS booklet contains 156 variables. The variables are subdivided into the following eight categories: administration variables ($n = 11$) pertaining to information regarding the police organization; victim variables ($n = 23$) such as the victim's height and build; offender variables ($n = 40$) such as the offender's height and build; vehicle variables ($n = 16$) pertaining to aspects of any vehicles involved in the crime; crime scene variables ($n = 9$) pertaining to details of the crime scene such as location; offence variables ($n = 48$) such as the nature of the offence (e.g., sexual); weapon variables ($n = 5$); and summary variables ($n = 4$) such as an open-ended question requiring them to write out a summary of the incident and space to provide any additional information pertinent to the offence.

Procedure

Each participant was tested individually at the participating police organization's headquarters. All participants were presented with both an informed consent form and a short verbal presentation on the purpose of the study and their role as a participant. The participants were informed that the goal of the study was to measure the data entry accuracy of police

officers. Participants were then given the case file to read at their desired pace and were asked to complete a ViCLAS booklet as they would normally do following one of their investigations. The participants were also told to be as accurate as possible when completing the booklet. After completing the booklet each participant was asked to complete an eight-item demographic questionnaire. On average, the participants took approximately 45 min to read the case file and 70 min to complete the ViCLAS booklet. To prevent the integrity of the study being compromised by having debriefed participants communicate about the study with future participants, a debriefing form outlining the true purpose of the study (i.e., reliability in coding) was emailed to the participants after all testing was completed.

Inter-rater Reliability Calculations

Percentage occurrence agreement was calculated using the method outlined by Hartmann (1977). For each variable, all possible pair-wise comparisons between participants were first established. The total number of instances where two participants in a pair agreed that a specific variable was present in the case file was then tallied. This value was then divided by the total number of pairs where at least one participant in the pair indicated that the variable was present. This proportion was then multiplied by 100 to arrive at a percentage occurrence agreement. This was the same procedure used by Martineau and Corey (2008).

For clarity, consider an example where four police officers provide the following yes or no decisions about the use of a knife in an offence: officer 1: yes, officer 2: no, officer 3: yes, officer 4: no. From this example, six pair-wise comparisons would emerge: yes-no (officer 1 vs. officer 2); yes-yes (officer 1 vs. officer 3); yes-no (officer 1 vs. officer 4); no-yes (officer 2 vs. officer 3); no-no (officer 2 vs. officer 4); yes-no (officer 3 vs. officer 4). The no-no pair-wise comparison would be removed as this represents a non-occurrence agreement (i.e., neither of the

police officers indicated that a knife was used in the offence). Of the five remaining pair-wise comparisons, there is one agreement and four disagreements, thus resulting in a percentage occurrence agreement of 20% (i.e., 1/5).

Percentage occurrence agreement was calculated for 106 of the 156 variables comprising the ViCLAS booklet. One of the variables contained in the summary section (i.e., a summary of the case material) was not included in the analysis because the information provided by this variable is redundant (i.e., provided elsewhere throughout the booklet) and is a replication of the materials provided in the experiment. In addition, as this study used percentage occurrence agreement, any time a participant answered 'unknown' or left a variable blank, their answers were not included in the calculations. There were 49 variables throughout the booklet that were left blank by all 10 participants and were omitted from the analysis.

Any of the 106 variables requiring an open-ended response were coded to identify the total number of discrete responses. Each unique response was then treated as a sub-variable, subjected to inter-rater reliability analysis, and collapsed to form the percentage occurrence agreement for that main variable. For example, a response to a question (i.e., main variable) regarding what the offender said to a victim (verbal exchange) that included "I am going to kill you," "You better listen to me," and "Keep quiet" would have been treated as three sub-variables. How often officers agreed on each of these phrases would be calculated and then combined for a total score for that main variable.

Similarly, every variable that had the option to "check all that apply" was also subjected to multi-level inter-rater reliability analysis, and collapsed to produce a single inter-reliability score for that main variable. For example, a question pertaining to whether the offender was at risk of being detected at the scene and had three options (e.g., area deserted, potential to see

offender, potential to hear/interrupt offender) would have been treated as three sub-variables.

How often officers agreed on each of these options would be calculated and then combined for a total score for that main variable.

Reliability of the coding was assessed by having an independent researcher code each of the booklets. The independent coder was provided with a 1-h training session that covered the practical aspects of coding the booklets, the structure and content of the coding guide, and the content dictionary. The reliability of coding was measured using Cohen's Kappa (Cohen, 1960). The Kappa value for the administration section was .86, .88 for the victim section, .98 for the offender section, 1.00 for the vehicle section, .98 for the scene section, .87 for the offence section, and 1.00 for both the weapon and summary sections. The average Kappa across all eight sections was .95, thus suggesting excellent agreement between the coders (Fleiss, 1981; Landis & Koch, 1977).

All analyses reported below are derived from 106 reliability scores. Across the 106 variables (including all sub-variables) there were a total of 8,863 pair-wise comparisons. In the current study, an 80% level of agreement was used as the minimum acceptable level of occurrence agreement (Hartmann, 1977).

Results

Across the 106 variables, the average overall percentage occurrence agreement was 30.77% ($SD = 30.23$, 95% $CI = 25.02, 36.53$). The levels of percentage occurrence agreement for each of the eight sections of the ViCLAS booklet are shown in Table 1. As can be seen, agreement ranged from a low of 2.36% for weapon variables to a high of 62.87% for administration variables. In addition, the width of the confidence intervals exceeded 20% for the administration, victim, offender, and scene sections and exceeded 10% for the vehicle and

offence sections. There was also variability within seven of the categories, with a range of 5.90% for weapon variables and a range of 100% for victim, offender, and offence variables.

The number of variables in each section that met or exceeded the acceptable level of agreement is also shown in Table 1. In total, 11 (10.38%) of the variables met the acceptable level of agreement. For the administration section, only incident type reached the acceptable benchmark. In the victim section, three of the 15 variables met the acceptable level of agreement. These three variables were the victim's status (i.e., they survived an attack), the victim's gender, and the victim's occupation. For the offender section, four of the 30 variables met the acceptable level of agreement. These four variables were the offender's gender, hair shade, hair colour, and language spoken. Two of the 9 variables comprising the scene section had an acceptable level of agreement. These variables were the offender's risk of being detected at the scene and a description of the scene (i.e., indoors or outdoors). For the offence section, only the variable pertaining to how the offence ended (i.e., release, escape, rescue, or death) reached an acceptable level of agreement. None of the variables for the vehicle, weapon, or summary sections exceeded the level of acceptable agreement.

Discussion

The purpose of this study was to conduct an independent test of inter-rater reliability associated with ViCLAS - a popular crime linkage analysis system that is used internationally for both investigative and research purposes. As predicted on the basis of previous research (Martineau & Corey, 2008), the level of occurrence agreement found in the current study was, for the vast majority of ViCLAS variables, unacceptably low. Even if one focuses on the upper limits of the CIs, which is the most optimistic estimates of the levels of reliability, the levels of agreement are still very low (with the exception of the administration section). Given that our

study dealt with some of the limitations of Martineau and Corey's study and still resulted in low levels of reliability, our findings raise serious concerns about the validity of inferences drawn using ViCLAS data and, potentially, the validity of inferences drawn from other linkage systems in use around the world. Similar concerns exist for researchers who are using ViCLAS data as the primary source of information for their research endeavours. Being able to rely on ViCLAS requires empirical evidence demonstrating the reliability of data contained in the system; evidence that is not yet available.

What about the 11 variables that exceeded acceptable levels of agreement? Although it is a positive sign that certain ViCLAS variables could be coded reliably, close inspection of these variables shows that they may be of little use for making linkage decisions. For example, some of the variables that participants agreed upon may be too common (e.g., the offender's gender) or too subjective/subject to change (e.g., the offender's hair colour). Moreover, many of the variables that one would intuitively think of as useful for establishing crime linkages (e.g., the sexual acts that occurred during the offence, whether or not the victim was targeted by the offender, the type of con or deception used) had very low levels of agreement (all < 25%). Understanding why these variables are associated with low levels of inter-rater reliability may be important for improving the reliability of ViCLAS coding and ultimately the effectiveness of this renowned linkage system.

Potential Explanations for the Low Levels of Inter-rater Reliability

There are several potential explanations for the low levels of agreement observed here. First, it is possible that participants may have viewed the task as boring and/or unimportant and, thus, put little effort into their coding of the case material. Social psychological research has provided evidence for both motivational and performance decreases when individuals view tasks

as mundane and inconsequential (Bartis, Szymanski, & Harkins, 1988; Healy, Kole, Buck-Gengler, & Bourne, 2004; Sheppard, 1993). If boredom with the coding task, or its perceived lack of importance, is causing motivation and/or performance decreases, then it will be important to ensure that this is not the case for coders using ViCLAS booklets in naturalistic settings. The ViCLAS booklet contains a large number of variables, takes time to complete, and officers are often not made aware of linkage successes (as our post-study questionnaire results suggest). These conditions may erode the extent to which investigators perceive the coding task as important and, as a consequence of less motivation, inter-rater reliability is likely to suffer. Examining how boredom, or perceived lack of importance, influences the reliability of ViCLAS coding is an important issue to be examined in future research.

Second, the lack of intimacy with the case file information might also explain the low level of agreement among our participants. Of course, participants would be likely to have more knowledge of cases that they have investigated. It is therefore possible that the increased level of intimacy with those cases could result in higher inter-rater reliability scores. That being said, the participants in this study had the opportunity to become as familiar with the case material as they wished. They were also given the opportunity to make notes, refer back to the file, and work at their own pace. Furthermore, it is conceivable that there are instances where police officers who are not associated with the investigation are asked to complete the ViCLAS booklets due to resource demands. Nevertheless, the effect of case familiarity on coding reliability is an empirical question that requires testing.

Third, inexperience with ViCLAS booklets may be another explanation for the low levels of inter-rater reliability. Some of the participants reported having little experience filling out ViCLAS booklets and minimal training on how to complete them. Potentially compounding this

issue is the fact that the participants were not provided with a field investigator's guide. Taken together, these two issues might have caused the participants to struggle with completing the various sections of the booklet. On the other hand, it should be noted that our findings are likely to have a high degree of external validity because our participants investigate ViCLAS-appropriate crimes and are in a position to complete ViCLAS booklets. The participants also rated the case file as being very similar to the materials typically available to them when completing ViCLAS booklets. Finally, the participants reported that they are unfamiliar with the field investigator's guide and do not tend to use it when completing ViCLAS booklets; thus, validating our decision not to provide it to them during the coding task. Furthermore, the level of reliability exhibited by the participants in Martineau and Corey's (2008) study (who were given the investigator's guide) was similar to the reliability exhibited by our participants (who were not given the investigator's guide). Examining how ViCLAS training, or experience in completing ViCLAS booklets, impacts the reliability of ViCLAS coding should be a priority in future research.

Fourth, the nature of the ViCLAS booklet itself may have caused the low levels of inter-rater reliability. For example, the sheer number of variables (and subsections) contained in the booklet may make it difficult for participants to obtain high levels of reliability – which is what officers contend with in reality. It may be the case that officers also struggle with the ambiguity and interpretation of some of the variables. For example, variables such as “area was essentially deserted” and “how would you rate the extent of offender/victim negotiation?” are highly subjective variables that are likely to lead to disagreements among officers. In addition, there are a number of seemingly complex and confusing questions that may cause problems for officers. For example, a few of the questions in the ViCLAS booklet require multiple forced choice

answers that provide an increased probability of obtaining disagreements. It is possible that the size and nature of the ViCLAS booklet may be cause for concern and research should continually be undertaken to determine if and how the ViCLAS booklet can be revised so as to increase its reliability.

Fifth, and perhaps related, is the role that the case material played in the reliability calculations. Regardless of participant knowledge or experience with ViCLAS, providing people with a great deal of crime-related information and then asking them to dissect and transfer that information into a detailed booklet (as done after actual investigations) would likely result in low levels of inter-rater agreement. Comprehension of the material is a possible concern, as is retention of the information if investigators rely on their memory of the case material (versus direct referencing) when completing the ViCLAS booklet.² It may also be the case that the information contained in our experimental material was particularly vague, convoluted, or complex. The reality, however, is that the case file used in the current study is a genuine case that officers investigated and had to enter into a ViCLAS booklet. Careful attention should nevertheless be paid to these issues when preparing experimental material for future studies.

Limitations of the Research

There are some potential limitations with this study. For example, it could be argued that our small sample size is a potential limitation. In response to this, we can simply say that the findings based on our 10 participants are consistent with Martineau and Corey's (2008) findings that were based on over 200 participants. Perhaps more importantly, however, is the fact that our results are based on nearly 9,000 pair-wise comparisons. The number of comparisons between participants, and the fact that the current results match the trends reported by Martineau and Corey (i.e., low reliability, variation in agreement across sections), should reduce any concerns

about the sample size. Moreover, in reality, finding disagreements between just two officers should be enough to raise concerns about the reliability of ViCLAS data.

Another potential limitation pertains to the reliability of our results given that we used only one case file (and one crime type). Our results could be due to something unique about our experimental material. Ideally, researchers would have participants in future studies complete multiple ViCLAS booklets using a range of case files. However, getting officers to complete this task will pose logistical challenges (e.g., with respect to the amount of time required). It is likely this issue is best resolved through a convergence of evidence, where inter-rater reliability is tested in a series of studies that use a single case file, but these case files would vary systematically (e.g., crime type, amount of information in the case file).

Some may also argue that our findings are not generalizable because of the artificial nature of our testing conditions. We disagree with this argument. If participants are unable to obtain high levels of inter-rater reliability under highly controlled experimental conditions, where they are not distracted or overburdened, then it is likely that coding ViCLAS booklets in actual police settings would result in even lower levels of reliability. For example, research shows that distractions have a negative impact on performance and accuracy in a range of complex tasks (Banbury, & Berry, 1998; Kemker, Stierwakt, LaPointe, & Heald, 2009; Sanders, & Baron, 1975). The crowded offices, background noise, telephone calls, and interruptions from colleagues that are a part of every investigator's work environment, are likely to have a detrimental rather than facilitatory effect on the reliability of genuine ViCLAS coding.

Conclusion

Regardless of whether crime linkage systems are used for operational or academic purposes, the ability to use these systems to make valid inferences is dependent on the reliability

of the data that are stored in them. Our results, and those of others (Martineau & Corey, 2008), suggest that the data contained in ViCLAS may be unreliable. Continued use of a system that may contain unreliable data can be difficult to justify, especially given the many serious consequences attached to linkage decisions. For example, pursuing potential linkages derived from unreliable data can be a waste of valuable police resources, taxpayer dollars, and may even result in individuals being improperly considered or even falsely accused of crimes they did not commit. Time spent on ViCLAS-related activities also takes investigators away from other important tasks that may be more pertinent.

As philosophers of science have argued for some time, if evidence for a technique's validity is largely negative or unknown, and the technique has the potential to cause harm, the appropriate stance is to advocate for prohibitions against the continued use of that technique until data supporting it emerges (Lilienfeld & Landfield, 2008). We anticipate that such an argument will be viewed negatively by some and will be met with counter-arguments that police organizations do not have the time to wait for research to be conducted as crimes continue to be committed. In response to such anticipated arguments, we simply state that without such research, there is a real risk that linkage systems will never reach their full potential and may cause more harm than good. Instead of viewing our findings in a negative light, we hope our findings act as an impetus (especially for researchers and practitioners who advocate such systems) to find ways to improve the reliability of data contained in linkage systems.

References

- Bennell, C., Snook, B., MacDonald, S., House, J., & Taylor, P. J. (2011). *Crime linkage systems: A critical review and research agenda*. Manuscript submitted for publication.
- Banbury, S., & Berry, D. C. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology, 89*, 499-517.
- Bartis, S., Szymanski, K., & Harkins, G. (1988). Evaluation and performance: A two-edged knife. *Personality and Social Psychology Bulletin, 14*, 242-251.
doi:10.1177/0146167288142003
- Bijleveld, C., & Smit, P. (2006). Homicide in the Netherlands: On the structuring of homicide typologies. *Homicide Studies, 10*, 195-219.
- Clark, D. (2002). *Dark paths, cold trails: How a Mountie led the quest to link serial killers to their victims*. Toronto, ON: Harper Collins.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Experiment, 20*, 37-46. doi:10.1177/001316446002000104
- Collins, P., Johnson, G., Choy, A., Davidson, K., & MacKay R. (1998). Advances in violent crime analysis and law enforcement: The Canadian Violent Crime Linkage Analysis System. *Journal of Government Information, 25*, 277-284. doi:10.1016/S1352-0237(98)00008-2
- Egger, S. (1984). A working definition of serial murder and the reduction of linkage blindness. *Journal of Police Science and Administration, 12*, 348-357.
- Fleiss, J. (1981) *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- Friesen, J. (2004, January 19). How police crack open criminal minds. *The Globe and Mail*. Retrieved October 29, 2009, from <http://www.theglobeandmail.com>

- Grubin, D., Kelly, P., & Brunson, 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
- 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
- Kemker, B. E., Stierwakt, J. A., LaPointe, L. L., & Heald, G. R. (2009). Effects of a cell phone conversation on cognitive processing performances. *Journal of the American Academy of Audiology, 20*(9), 582-588. doi:10.3766/jaaa.20.9.6
- Kocsis, R., Cooksey, R., & Irwin, H. (2002). Psychological profiling of sexual murders: An empirical model. *International Journal of Offender Therapy and Comparative Criminology, 46*(5), 532-554. doi:10.1177/030662402236739
- 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*(3), 183-191. doi:10.1002/jip.51
- Landis, J. B., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- Lilienfeld, S. & Landfield, K. (2008). Science and pseudoscience in Law enforcement: A user-friendly primer. *Criminal Justice and Behavior, 35*, 1215-1230.
- 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(2), 241-253. doi:10.1007/s10508-005-1801-2
- 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(4), 384-406. doi:10.1177/0306624X06298465
- Ormerod, D., & Sturman, J. (2005). Working with the courts: advice for expert witnesses. *The forensic psychologist's casebook: Psychological profiling and criminal investigation* (pp. 170-193). Devon United Kingdom: Willan Publishing.
- Royal Canadian Mounted Police (RCMP) (n.d.). Violent Crime Linkage System (ViCLAS). Retrieved from <http://www.rcmp-grc.gc.ca/viclas-salvac/index-eng.htm>
- Sanders, G. S., & Baron, R. S. (1975). The motivating effects of distraction on task performance. *Journal of Personality and Social Psychology*, 32(6), 956-963. doi:10.1037//0022-3514.32.6.956
- Sheppard, J. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin*, 113, 67-81. doi:10.1037//0033-2909.113.1.67
- State v. Code*. (1994). 627 So.2d 1373.
- State v. Fortin*. (2004). 178 N.J. 540: 843 A.2d 974
- State v. Prince*. (1992) 9 CAL.APP.4th 1176, 10 CAL.RPTR.2D 855.
- State v. Pennell*. (1989). Del.Super., 584 A.2d 513.

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

doi:10.1348/135532506X118631

Table 1

The Total Number of Variables Tested, Mean Percentage Occurrence Agreement, 95% Confidence Interval, Range of Percentage Occurrence Agreement, and the Number of Variables Meeting the Acceptable 80% Agreement Criteria.

Booklet Section	Number of Variables Tested	Mean Percentage Occurrence Agreement	95% Confidence Interval	Range of Percentage Occurrence Agreement	Number of Variables Meeting 80% Criteria
Administration	5	62.87%	34.24% to 91.51%	19.34% – 100%	1
Victim	14	40.86%	21.39% to 60.32%	0% - 100%	3
Offender	30	37.93%	26.95% to 48.91%	0% - 100%	4
Vehicle	6	11.21%	5.21% to 17.20%	2.95% - 20%	0
Scene	9	40.19%	16.56% to 63.82%	4.17% - 100%	2
Offence	36	21.78%	14.55% to 29.01%	0% - 100%	1
Weapon	5	2.36%	0% to 5.19%	0% - 5.90%	0
Summary	1	12.50%	--	--	0
Overall	106	30.77%	25.02% to 36.53%	0% - 100%	11

Note: The average of the mean percentage occurrence agreements for the eight sections is 28.71% (95% CI = 14.75%, 42.68%).

Footnotes

¹ Martineau is the Manager of Research and Development for the Behavioural Sciences Research Branch of the Royal Canadian Mounted Police – the developers of ViCLAS.

² It is interesting to note that many of the participants reported, in passing, that they often rely upon their memory (rather than making direct reference to the case materials) when completing ViCLAS booklets.

Author Biography

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.

Kirk Luther is a graduate student at Memorial University of Newfoundland. His research interests include the cognitive processes underlying judgment and decision making, crime linkage systems, and police interviewing and interrogation methods. His other research interests include distinguishing science and pseudoscience in law enforcement.

John C. House is a Superintendent of the Royal Newfoundland Constabulary (RNC) in Canada. He holds a Master of Science Degree 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.

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.

Paul J. Taylor, PhD, CPsychol, 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 behavioural science and

technology to address areas such as protective security, vulnerable witness interviewing, and the investigation of terrorism.