Column: CCOPP's Stories*
(*Canadian Committee of Police Psychologists)

By Craig Bennell, Ph.D.,
Director-at-Large: Police Psychology

A Call for the Use of Statistical Prediction Rules in Police Investigations

Police investigators routinely make important decisions, many of which affect people's lives. An investigator might have to determine whether a suspect is lying, whether a suicide note is genuine, whether a bite mark was made by an adult, etc. Two-alternative decisions of this type are referred to as diagnostic tasks in most settings, and across a range of fields, including radiology, engineering, and psychology, they are now often dealt with through the use of statistical prediction rules (SPRs). However, despite the fact that many researchers and practitioners are becoming more familiar with these tools, they are all but unknown in the policing domain. This is unfortunate given the bulk of empirical evidence, which tends to support their use.

Importantly, this does not need to be the case. The same goals and problems exist in each of the above fields and SPRs can be as useful in policing as they are every where else. Ultimately, the goals in all of these areas are to increase the accuracy and utility of the decisions that are made. The problems are to identify the best predictor variables for a given task and to set an appropriate decision threshold for determining when an event of interest has occurred (or will occur in the future). These are problems, of course, because typically there will be many predictor variables from which to choose (with only some being useful) and rarely will it be obvious where a decision threshold should be placed.

Enter SPRs and, of equal importance, a method for evaluating the predictive accuracy of these rules. Given the abundance of statistical methods now available, the choice of a prediction tool is largely a matter of preference (e.g., one could use a regression model, a neural network, a genetic algorithm, etc.). However, more and more often, receiver operating characteristic (ROC) analysis is being viewed as the technique of choice for evaluating these tools. Indeed, in the most recent issue of *Law and Human Behaviour*, two of our colleagues, Marnie Rice and Grant Harris, recommended that this procedure be the standard method for measuring diagnostic accuracy in forensic psychology.

As most of you will know, much of the value of ROC analysis comes from the fact that it can provide a meaningful measure of diagnostic accuracy in the form of the area under a ROC curve (AUC). Unlike most measures of accuracy, such as the percentage correct, the AUC does not depend on the arbitrary selection of a decision threshold. Instead, it reflects the position of an entire ROC curve in its graph, constructed by examining the relative frequencies of all possible decision outcomes for a given task across many thresholds. The AUC, therefore, is a more valid indicator of decision-making performance. But what does this mean for the police decision-maker?

Basically, what it means is that ROC analysis can be used to examine issues arising in the policing context in a way that is more valid compared to what is currently done. In my own research in the area of police decision-making, I have encountered at least five policing issues that would benefit greatly from the use of ROC analysis. It could be used to establish the predictive accuracy of a specific diagnostic system, identify the most accurate predictors for a given diagnostic task, set decision thresholds in order to maximize decision-making utility, compare the ability of different decision-makers, or examine the impact of situational factors on decision-making performance.

To demonstrate the value of SPRs generally, and ROC analysis specifically, consider the common investigative task of deciding whether two crimes have been committed by the same offender. In the absence of forensic evidence, behavioural evidence must be relied on to complete this task. This can include an analysis of information related to the type of victim that was selected, the time of day when the offence took place, the geographic location(s) of the attack, behaviours exhibited at the crime scene, and so on. The question for the police investigator is, given two crimes, which piece(s) of information will result in good decisions.

What the investigator has to do is identify behaviours that are likely to be repeated across crimes committed by the same offender, but these behaviours cannot be exhibited by all other offenders. In our studies of this task, which have focused solely on burglary, we have found that police personnel perform at levels approximating an AUC of .60. The reason our participants do not perform better than this is that they tend to rely on behaviours that are low in predictive power. For example, it is rare for individuals not to focus on the type of property stolen when establishing links (e.g., they will indicate that cash was taken in both crimes). Yet, in our samples, cash is nearly always the only item that offenders steal.

Such a scenario would suggest that SPRs may be of some use. And it turns out that they are. For example, when examining the SPRs we have developed for the purpose of linkage analysis, we often achieve AUCs in excess of .80, even when the SPRs are applied to new samples. Why do these SPRs perform better than police personnel? For one reason, they rely on variables that consistently outperform (in

terms of their predictive power) the behaviours that are focused on by the individuals we have tested.

By far, the best linking variable relates to the distance between offence locations, what we refer to as the inter-crime distance. It turns out that the closer two offences are, the more likely they are to have been committed by the same offender. It is in fact this one variable that accounts for the superior performance of our SPR. This variable is rarely, if ever, mentioned by our participants. Fortunately, we have recently discovered that this SPR can be used to improve the performance of police personnel. For example, simply by telling these individuals about our findings, we are able to increase their accuracy significantly, to about .70. However, our participants still hold on to many of their existing beliefs, and rarely do they perform as well as the SPR.

The fact that inter-crime distance is a good linking variable brings me to the second advantage of ROC analysis (the first advantage being that it provides a valid method for comparing different decision-makers [SPRs and humans] on a given task). The second advantage relates to the use of ROC analysis for setting thresholds. Given that inter-crime distance is an effective linking variable, this begs the question: how close do two crimes have to be to one another before we should decide that they are linked?

The answer to this question turns out to be crucial. For example, in one of our studies, the ratio of hits to false alarms varied drastically when we shifted the threshold by only a few kilometres. At a threshold of 3 km, the hit rate was .91, whereas the false alarm rate was .50. At a threshold of 1 km, the hit rate was .72, whereas the false alarm rate was only 0.08. These are not small differences. They indicate that, simply by adjusting the threshold slightly, one could massively decrease the likelihood of making a false alarm (saying two crimes were linked when they aren't) while reducing the chance of making a hit (saying two crimes are linked when they are) by a much smaller amount. The optimal threshold for this sample ended up being about 2.5 km. Although we have yet to do so, providing this additional information to individuals would likely increase their ability to make good decisions, perhaps putting them on par with our SPR.

Although I have focused on one investigative task, the same argument applies to many others (indeed, research in my lab has shown that the above findings mirror what happens with many other investigative tasks). The value of SPRs has been shown in other areas and I see no reason why the same would not be true in policing. Given the recent creation of our sub-section on police psychology, and a great newsgroup to go along with it, it seems a perfect opportunity to begin discussing these issues. Indeed, the mix of practitioners and researchers that make up this group is what is needed to make this happen. Police psychology is certainly lagging behind other sub-disciplines of forensic psychology. Addressing the issue raised in this article may be one way to start catching up. All comments are welcome.

Column: Beyond a Reasonable Doubt

By Joanna Pozzulo, Ph.D.
Director-at-Large: Psychology in the Courts

The Bug Effect

I recently started my subscription to University Affairs. When my first copy arrived (December, 2005), I was quite intrigued to read the headline, "The CSI effect" (McCabe, 2005), thinking the article would have something to do with crime shows and their influence on the general public's awareness of forensic issues. I turned the page only to see the caption, "*The old days of eyewitness testimony in criminal trials are largely over".* What luck I thought, a perfect topic for debate.

Has the 'criminal trial' become one where only forensic anthropologists, entomologists and other "forensic science" types are welcome? Is there no place for eyewitness testimony? Or testimony from eyewitness experts? Are psychologists who research forensic issues not considered "forensic scientists"? Are they less of a scientist than forensic entomologists?

A quote in the University Affairs article by Dr. Gail Anderson, one of relatively few Canadian forensic entomologists, states, *"Witnesses can have poor memories, bad eyesight and questionable motives. Physical evidence, if it is done correctly, does not lie."* (p. 20). I pondered her statement for a while, then, wondered how knowledgeable are other scientists with regards to psychological science in general and eyewitness research in particular? What if eyewitness evidence is done correctly?

Yes, eyewitnesses can make mistakes. We have seen erroneous eyewitness identification be the primary evidence convicting innocent defendants (Wells et al., 1998). Some of these innocent defendants have spent 10, 15, even 20 years or more incarcerated. But, and this is a big but, eyewitnesses are also correct. Witnesses can describe what happened. Witnesses can describe what the culprit looked like. Witnesses can correctly identify the culprit. Across several research studies, it is not uncommon to see correct identification rates at 75% or higher (when the culprit is in the lineup presented, e.g., Leippe, Romanczyk, & Manion, 1991). Moreover, just as there are procedures that can make the collection of physical evidence more reliable, there are procedures that can make the collection of eyewitness