Chapter 1:
From Mysticism to Science: An Overview of Risk Prediction

David Riley

When the dying Caesar eyed his assassin and rued the fact that he had paid scant attention to Cassius's "lean and hungry look," he exemplified that very human tendency to attribute qualities and intentions to others on the basis of quite irrelevant characteristics. Drawing parallels between certain characteristics and what a person might do, or crude stereotyping on the basis of ethnicity or religious belief have formed the basis for subjective judgments throughout recorded history, and this has been re-enforced in literature from Sophocles to T.S. Eliot. Similarly, the accumulation of each depravity and evil deed has also been seen as having a sequel in physical presentation as was the case in Wilde's *Portrait of Dorian Gray*, or Zola's *La Bête Humaine*.

Such approaches were formalised by investigators in the 19th Century such as Gall, the Swiss phrenologist who persuasively argued that character and temperament could be accurately inferred from the shape of the skull, and Lombroso who went further in associating a wide variety of attributes with bodily characteristics. While such notions strike the modern student of social science as fanciful and devoid of credibility, it almost seems part of the human condition to base judgments about people, (and implicitly about what they will do in the future) on their appearance and manner of relating.

Notwithstanding the vast body of research which attests to the fallibility of judgments based on observation and impression, even those individuals who should (and indeed do) know better persist in basing important decisions about things such as appointments to positions of responsibility or release of offenders or mental patients into the community, on impressions rather than well established principles of risk assessment.

Tarasoff

The world changed for those charged with the provision of social services in 1976 when the California Supreme Court handed down its landmark decision (*Tarasoff v Regents of the University of California*) which clearly signalled a liability on the part of professional psychologists in relation to the behaviour of their clients. The *Tarasoff* decision was a result of legal action

following the murder of a student, Tatiana Tarasoff, at the Berkley campus of the University of California by a graduate student from India, Prosengit Poddar. The background to this litigation is both fascinating and salutary.

Poddar had become infatuated with Miss Tarasoff following her paying him some attention at a New Years Eve party at the end of 1968, but he became depressed when she subsequently made it clear to him that his attentions were neither welcomed nor reciprocated. At the urging of a friend, Poddar attended the student health facility on campus where a psychiatrist referred him to a psychologist for psychotherapy. Poddar discontinued his treatment, but because he had disclosed to the therapist that he had intentions of killing Miss Tarasoff, and that he had purchased a weapon for this purpose, the psychologist and two supervising psychiatrists agreed that the patient should be hospitalised involuntarily for evaluation, and the campus police were asked to detain him for that purpose.

When the campus police interviewed Poddar, he reportedly appeared rational and coherent, and on that basis they accepted his reassurances that he would stay away from Miss Tarasoff and he was released. Two months later, Poddar murdered Miss Tarasoff. The Tarasoff family sought legal redress and the California Supreme Court subsequently found in favour of the plaintiff in that the psychologist did not carry out his duty to warn the likely victim. This decision noted that "once a therapist does in fact determine, or under applicable professional standards reasonably should have determined, that a patient poses a serious danger of violence to others, he bears a duty to exercise reasonable care to protect the foreseeable victim of that danger". Subsequent Court rulings in North America have reaffirmed this general principle, and have broadened the scope of *Tarasoff* to include those closely related to potential victims, and more recently there have been successful claims brought on the basis of property damage.

While the Tarasoff case clearly signalled an obligation on the part of mental health professionals to fulfil a duty of care, those charged with such a responsibility could have taken little comfort from the results of two studies which were published in the mid 1970s. By an unhappy coincidence, the *Tarasoff* decision was thrown into sharp relief by the results of these two "natural experiments" which resulted from court decisions brought under American civil rights legislation.

In the first of these, a 1966 Supreme Court decision in the case of Johnny Baxtrom (*Baxtrom v Herold*), the Court ruled that Mr Baxtrom's continued retention in a civilian mental health facility following his discharge from a prison hospital for the mentally ill inmate was unconstitutional. Such retention of mentally ill inmates following the expiry of their sentences was not uncommon in New York State, and as a result of the *Baxtrom* decision,

nearly 1,000 inmates in hospitals for the criminally insane were transferred to civil facilities, and eventually about half of these were released into the community. This allowed for the kind of natural experiment which could never have been planned and conducted by social scientists. Steadman and Cocozza (1974), were thus able to test the predictions of dangerousness which had resulted in these patients being confined within mental health facilities.

These researchers were able to follow up 98 patients released from civil confinement, and they found that only two variables had at least some relationship to arrests during a four-and-a-half year follow up period; these being age and previous criminal history score. Not surprisingly, youthfulness and extensive histories of offending were associated with a higher likelihood of being arrested. The predictive power of these two measures was, however, modest; the best result which they could achieve in identifying 17 of these patients as high risk was a corresponding misclarification of 19 who were not arrested following their discharge. Steadman and Cocozza (1974) concluded, "any enthusiasm for success is tempered by the knowledge that these measures at best…would still mean that any special programme of preventive detention would for every 100 patients classified be inaccurate for over half of them".

Steadman and Cocozza (1974) also attempted to determine the accuracy of predictions of violence. Of a total of 20 arrests, only seven had been for violent offences. An evaluation of hospital records indicated that a further seven individuals in the sample had been hospitalised as a consequence of violence or dangerous behaviour. Applying their prediction device to these fourteen individuals resulted in predictions which correctly identified eleven of those who subsequently exhibited violent behaviour but these measures also predicted a further 25 individuals as being in that category. In his forward to Steadman and Cocozza's (1974) book, *Careers of the Criminally Insane*, Toby described the results "…we have used a meat axe to kill a spider". In short, the high preponderance of false positive errors demonstrated that clinicians and administrators assumed patients to be far more dangerous than they actually proved to be.

Social scientists were able to get a second bite of the prediction cherry following a further Court decision in Pennsylvania in 1969. In *Dixon v Attorney General of the Commonwealth of Pennsylvania*, part of the Pennsylvania Mental Health and Mental Retardation Act, was ruled unconstitional, which resulted in the transfer of 586 male patients at Fairview Hospital for the criminally insane to civil institutions. Many of these were subsequently discharged into the community and Thornberry and Jacoby (1979) set out to replicate the findings of the *Baxtrom* study with some methodological

refinements. They followed up 414 patients for an average of 3 years. Of these, 24% were subsequently arrested on at least one occasion. This finding corresponded to the *Baxtrom* rate of 20.4% and on that basis the authors concluded that the assessment made by the Fairview Hospital staff that these patients as a group were dominated by likely recidivists was not confirmed. Of those who were arrested, over one third were for property offences, and another quarter were for victimless public order offences. The authors followed the Steadman and Cocozza (1974) method of searching hospital files and these data, combined with those who had reoffended in a violent manner, resulted in 60 of the 414 releases being able to be classified as violent recidivists. Thornberry and Jacoby (1979) were unable to find variables, either singly or in combinations, that would yield a high predictive accuracy of violence. As in the *Baxtrom* findings, age was the most powerful single variable in predicting further violence, but the best which they could achieve combining age with the same dangerousness scoring method used by Steadman and Cocozza (1974) was an over prediction of violence at a rate of four to one.

Reviewing these findings in 1981, John Monahan emphasised the difficulties of low base rates and high false positive error, and in summarising the literature up to that time, drew to the attention of a responsive readership a number of other conceptual and practical problems in research on the prediction of violence. However, after highlighting all the imperfections in this area he concluded that, "careful predictions by some mental health and other workers are not only responsibly offered, but in some cases are vitally necessary".

The Inherent Dilemma in Predicting Risk

Those working in the provision of social services are constantly faced with making decisions which have to do with the potential of their clients to cause damage to others or harm themselves. Herein lies a dilemma, because predicting that a person may be dangerous to themselves or others can lead to that person's liberty being constrained in some way, be it by retention in prison, or in some form of mental health facility. The problem is that prediction is prone to error, and the rights of the individual must always be weighed against the potential consequences of their behaviour on others.

What the research stemming from the *Dixon* and *Baxstrom* court decisions indicated was that, while there was *some* ability to identify the majority of those individuals who might behave violently in the future, to achieve this resulted in very substantial rates of misclassification. In other words, for each

successful prediction, approximately three individuals were misclassified as being potentially violent when they were not. This is the inherent dilemma of risk prediction; to the extent that one is conservative with regard to the safety of the public, then the inevitable consequence is that greater numbers of people are misclassified with the potential curtailment of their freedom. Misclassification of people as high risk when they do not behave dangerously is called false positive error.

To the extent that the practitioner is concerned with the rights of the individual, and is very conservative about whom they identify as dangerous, then this will result in a number of people who go on to behave in a dangerous manner being misclassified. This type of misclassification is called false negative error.

Clearly, the rate of false positive error will depend upon the magnitude of the risk posed in any given situation, and if the potential is for serious negative consequences to impact upon large numbers of individuals, or if a particular behaviour is considered to be potentially lethal, then the tendency is to be more conservative in any risk prediction and the rate of false positive error will increase. Conversely, if the behaviour in question is considered to be less serious, and if the practitioner is concerned about the rights of the individual, then judgments of risk are likely to be less stringent, and the number of false positive errors will decrease, but with the inevitable consequence that the rate of false negative error will grow proportionately.

Types of Risk Assessment

While there are a number of approaches which are used in determining an estimate as to the risk that any given individual will behave in a certain way, they all assume a relationship between some aspects of the individual and future behaviour. The characteristics of the individual which may be employed to predict risk range from basic social and demographic characteristics such as gender, age, and occupation, aspects of past behaviour such as violent actions or past suicide attempts, descriptions of current levels of functioning such as employment stability or relationship status, through to clinical assessments which may also rely on diagnoses or psychometric measures of personality.

What all of these approaches have in common is the assumption that certain factors relating to the individual, be they a matter of historical record, a personality trait, or an appraisal of current functioning, bear a statistical relationship to what that person will do in the future.

Actuarial measures of risk

Actuarial measures of risk are the simplest and most basic risk assessment tools. They usually contain a list of factors which are checked as to their presence or absence. Sometimes the values assigned to each of these factors may be weighted according to the degree to which they are perceived to be associated with the event to be predicted.

One of the first systematic investigations which led to this type of scale was in the area of the prediction of recidivism, and was carried out by Ernest Burgess and was published in 1928. Burgess reported on the results of a study of more than 3,000 men paroled from a penitentiary in Illinois. Using their criminal records, Burgess coded 21 "facts" – type of offence, length of sentence, age, etc. – and then examined the relationship between these "facts" and subsequent parole outcomes. Not surprisingly, some variables showed a positive relationship to reoffending and a list of these variables provided the basis for the first objective prediction scale.

Such simplistic devices, while they certainly have their use and are an advance over unstructured approaches, are subject to obvious limitations. They are frequently statistically unsophisticated, merely relying on a demonstrated relationship between factors and some predetermined outcome; the problem with that approach being that sometimes several items are very highly correlated with each other and in a statistical sense contribute little unique variance to the scale's predictive accuracy. The danger here is that a single underlying trait or characteristic may in fact be picked up by several items, and disproportionately influence the level of risk which the scale purports to assess.

An example of this would be the use of age in the prediction of violence. If we analysed a large body of data on violent offenders, and correlated those data with future violent behaviour, we would find a strong statistical relationship between younger age and further violence. We would also be likely to find statistical relationships with a lack of trade qualification, never having been in a relationship in the nature of marriage for two years or more, frequent changes of accommodation, and higher levels of unemployment. All of these latter factors, however, would also be correlated with age, and by giving equal weight to all of those variables and including them in a prediction measure, we may actually decrease the predictive accuracy of the measure as we would (inadvertently) be biasing the measure unduly in terms of predicting higher rates of violence among the young than was warranted. This situation is known as multi-colinearity, and relates to a situation where including several variables which are highly correlated with each other may decrease rather than increase the performance of a risk scale.

A further problem with actuarial measures is that they typically focus on historical or unchangeable attributes. A person may never change their offending history, the number of past suicide attempts, or the sentences of imprisonments which they have received. Similarly, gender, ethnicity, and age are not within the individual's control, and prediction devices which rely on measures such as these are not subject to change, either as a result of efforts which people may have made to rehabilitate themselves, or overcome the difficulties which had previously contributed to their problems. To the extent that, for example, an assessment of future dangerousness rests on past episodes of violence and history of imprisonment, an individual's risk may never change, notwithstanding their demonstrating decades of socially constructive participation in the wider community.

Measures Based on Assessments of Dynamic Variables

While documented past behaviour and social and demographic characteristics show a statistical relationship to future events, they do not take account of individualised variables which may be changeable as a result of efforts made by people to overcome their difficulties or rehabilitate themselves.

For that reason, while risk scales based on actuarial measures are extremely useful and permit accurate screening of groups, a comprehensive assessment of risk should always take into account individual factors specific to the person being assessed. For example, in the prediction of deviant behaviour factors such as a person's ability to regulate their emotions, drug and alcohol abuse, employment stability, socially deviant peers, stable interpersonal relationships, and constructive use of leisure time all bear a relationship to reoffending. While there is still some debate as to the relative abilities of scales which rest on static variables and those which rely only on changeable factors to predict further criminal behaviour (e.g. Gendreau *et al*, 1996), a comprehensive risk assessment should clearly reference both sets of variables. In one sense, those static and unchangeable historical factors and immutable personal characteristics may best be seen as assigning the individual to a group with a known level of risk, while the clinical appraisal of dynamic factors related to the person's functioning serve to particularise the assessment to the individual, and allow for modification of the original baseline risk suggested by the static factors.

One further set of dynamic variables require mention here, and that is those which relate to stable personality traits which may be resistant to change. One of the most important of these which has received recent attention is psychopathy and its relationship to violence and exploitative behaviour. Once a reliable measure of this trait had been developed (Hare,

1990), a good deal of research has focused on the relationship between the interpersonal/ affective, and behavioural dimensions of this trait and various types of criminal behaviour. While still controversial, and arguably still not proven to be subject to modification by treatment, this personality variable may add to risk assessments undertaken by suitably qualified clinicians.

The Accuracy of Risk Assessment Measures

When a professional makes a judgment about a person's risk, whether it be their future risk to society or of harming themselves, that judgment has the potential to impact on the individual and may result in curtailment of their liberty in some way.

Clearly, then, it is important to know in advance the accuracy of the measure to be used, and to be able to accurately describe its strength and limitations to both those who are concerned with making decisions about the person's future, and to the individuals themselves. Unfortunately, both the ways in which risk measures are evaluated, and the way in which their accuracy is described may confuse rather than clarify.

Different risk measures may use various outcome measures; in the criminal justice area for example these measures may include parole failure, reoffending, reimprisonment, the commission of certain types of crime (violence or sex), or seriousness of offending. Also, the period over which future events may be predicted varies, from an immediate risk such as imminent violence or self harm, to risk over the longer term which may be months or years. Such differences in outcome measures, coupled with different time frames, means that it is often difficult to compare the accuracy of various measures, particularly as the settings in which they have been applied may also be different.

The situation is confused still further by different measures which purport to describe the accuracy of a risk assessment process. A cursory sampling of research articles revealed the following measures of risk scale performance; simple expressions of correlation, the common language effect size, area under the receiver-operating-characteristic curve, relative improvement over chance, and positive predictive power. Each of these measures describes the relationship of the scale or prediction instrument to outcome, but they all describe a different aspect of that relationship.

The two most commonly used measures which are applied to prediction scales are a measure of correlation, and the area under a receiver-operating-characteristic curve.

Correlation is a measure of association, and is expressed as a decimal between 0.0 and 1.0. A correlation of 1.0 would represent a total correspondence

between a measure and outcome whereas a value of 0.0 would indicate that there was no association between the measure and outcome to be predicted. Typically, provided the base rates of the behaviour to be predicted are not too low, the decimal value (e.g. 0.25) may be interpreted as a percentage measure of the association between the measure and the outcome (e.g. 25%). Typically, correlations of 0.3 are regarded as useful whereas correlations over 0.4 are regarded as indicative of excellent predictive performance.

The other widely used measure, the area under a receiver-operating-charactistic curve, derives from electrical engineering and signal detection theory, and arose out of attempts to describe the accuracy of detecting radio signals. The "area under the curve" (AUC) represents the relationship between correct predictions ("hits"), and incorrect predictions ("misses"). The area under the curve is perhaps the best measure for describing the overall capacity of any prediction measure, as it is not affected by low or high base rates of the behaviour to be predicted, and it captures more completely the performance of any measure. What the value of the area under the curve, which ranges from 0.5 which equates to chance alone, to (potentially) 1.0 which would be perfect prediction, is the proportion of times that an individual selected at random from a population exhibiting the behaviour which has been predicted would have a higher score than an individual picked at random from the population which did not exhibit the behaviour which was to be predicted. In other words, the area under the curve represents the sum of all possible pairs of comparisons of the measure when applied to a population with a known outcome.

Typically, areas under the curve of 0.65 and above are regarded as useful. Above 0.70 are considered to be good, and anything over 0.80 is regarded as excellent.

Whatever measure is used to describe the performance of a prediction model, there is one uniformly consistent finding in the field of risk assessment research. That is, that risk assessments which are made with the assistance of standardised risk assessment instruments universally outperform the unstructured clinical assessments of experts, be they psychologists, social workers, psychiatrists, or experienced custodial or hospital staff. This holds true when the judgments of "experts" are compared against even rudimentary risk scales comprising simple checklists of a few items.

Minimum Requirements for Risk Prediction

While the above does suggest that the latter part of the 20th Century saw a great deal of research on risk prediction of various kinds, and increasing sophistication in the methodology employed in risk assessment, it is still not

possible to provide the practitioner with a simple "recipe" as to how to go about it. There are, however, a number of minimum requirements which should apply to risk assessments, as those charged with the responsibility for making such assessments have a duty not only to the individual in question, but to the wider community. The following are offered as a list of minimum requirements for a competent risk assessment process; not necessarily representing a prescription of sufficiency for a professional risk assessment.

Use of a Structured Approach

The practitioner should apply a structured approach to the assessment, accessing and evaluating material related to a variety of predetermined areas. These should include a list of those factors known to be associated with the phenomenon to be predicted, and should be determined and itemised prior to the assessment interview taking place.

Access All Reasonably Available Information

Depending on the setting and the characteristics of the individual, often more credence may be placed on collateral and file information than on self reports. Particularly in forensic settings, individuals may have spent many years developing plausible rationalisations for their behaviour, and may present as experts in impression management. In such circumstances considerable weight should be given to reports of staff who may have had regular interactions with the individual, and also on other documentation which has accumulated over a period of years. Sometimes, file material is not always in one place, and if some documentation is reasonably available, attempts should be made to access such information in order that the most balanced and complete assessment of relevant risk factors can be made.

Be Aware of Base Rates

Base rates are absolutely critical when estimating levels of risk. Some behaviour, for example, reoffending following a prison sentence, may be more probable than not for the population being assessed. Other behaviour, such as serious self harm or sexual offending, may occur at a much lower frequency and over a much longer period of time. It is most important when making assessment of risk that the practitioner has a knowledge of base rates as they apply to the specific population in question, and is able to frame the assessment in terms of the average risk posed by a member of that group. For example, sexual reoffending has a comparatively low base rate (14%) over an extended (10 year) time period. Thus, for example, risk assessments should

be considered against that residual level of risk, and this should be reflected in any final professional opinion which is proffered.

Draw on Multiple Sources of Information

The literature endorses a multi modal approach to risk assessment. What that means is that competent risk assessments should involve a consideration of historical and static variables, assessments of current functioning, and also the influence of personality and temperamental factors. It is only when all three of these areas are considered that it is possible to provide a balanced assessment of the risk which the individual may pose.

Report the Positives as Well as the Negatives

When undertaking risk assessments it is very easy to focus on deficits and report solely on matters which contribute to heightened risk. In every situation, however, there are competing sets of factors which may be considered as "protective" and balanced and impartial assessment of risk should pay equal attention to the strengths of the individual, and those social and other factors which may mitigate risk, such as supportive family relationships, therapeutic progress which may have been made, or positive changes which may have occurred in the individual's circumstances.

Risk Assessments are Conditional

It is important that any risk assessment note that it is valid at the time which it is made and is dependent upon the individual's social and personal circumstances at that time. Reference should be made to future matters which could exacerbate or potentiate risk, including warning signs, and also note those matters which are within the individual's control which may, if addressed appropriately, serve to reduce the level of ongoing risk.

Do Not Oversell Risk Assessment

As has already been noted at some length, all risk assessment is prone to error, and whether that error is in the direction of over including individuals in a high risk category, or misclassifying larger numbers of high risk offenders as at lower risk, is dependent on the nature of the risk to be predicted. It is important, however, that all risk assessments clearly spell out the limits of accuracy inherent in the assessment, in order that those charged with making decisions about the individual have as much information as possible to inform any future course of action.

Beware of Over Estimating the Degree of Change

Inevitably, most practitioners put in the position of having to make assessments about a person's future risk will be specialists in the area of service provision. Such people are maintained in the helping professions by an inherent belief in the capacity of the individual to make changes and rise above their difficulties, and it is very easy for the treating professional to overestimate gains which the client may make in therapy. While outcome research in a number of areas of service provision clearly attests to its effectiveness, effect sizes for the most part remain relatively modest. What that means for the majority of clients is, of course, that the degree of reduction in risk as a result of successful participation in treatment programmes, even for those who have evinced most enthusiasm for the therapeutic process, is likely to be modest. For that reason, when service providers are also placed in the position of undertaking risk assessments in relation to their clients, the degree of change and implicitly the level of reduction in risk should always be the subject of discussion with a supervisor or colleague before any final opinion is provided.

In Conclusion

The area of risk assessment is complex and difficult. Although considerable progress has been made over recent decades, risk assessment is still fraught with error, and the practitioner must continually balance often conflicting information obtained from a variety of sources. While prediction is less than perfect, the results of such assessments have the potential to impact significantly on the lives of clients. There is little doubt that there will be incremental advances in technology in coming years, but decisions need to still be made currently on a daily basis about the level of restriction which should be imposed on some people, either for their own benefit or for the protection of the community. The practitioner is therefore faced with carrying out the task of risk assessment under less than ideal conditions, and with often inadequate background information. Nevertheless, such assessments are necessary, and despite their imperfections indubitably more professional and valid than in past decades.

The developments in this field are perhaps no more clearly exemplified than in John Mohahan's 1993 *American Psychologist* article on "Avoiding Therapist Exposure to *Tarasoff* Liability", in which he wryly observed that, following the publication of his seminal volume in 1981 on the prediction of violent behaviour, he had received many requests to appear

in court cases as an expert witness for the practitioner in matters to do with risk assessment. A dozen years on, however, those requests for his expertise in court continued to be made, but now by those claiming damages as a result of poor risk assessment by professionals!