

Final version prior to publication\_Ansbro, M. (2010) The nuts and bolts of risk assessment; When the clinical and actuarial disagree *Howard Journal of Criminal Justice* 49(3) 252-268

## THE NUTS AND BOLTS OF RISK ASSESSMENT: WHEN THE CLINICAL AND ACTUARIAL CONFLICT

MARIA ANSBRO

Senior Lecturer, University of Hertfordshire

Abstract: The aim of this research was to examine the thought processes that practitioners follow when they are conducting risk assessments. Weighing up the probability that an individual will inflict harm requires the practitioner to apply clinical and actuarial approaches, and integrate static and dynamic information. This is a complex and inexact task, and one that has been found lacking in reviews of serious further offences. This research focused on a small, atypical subgroup of risk assessments; those where the actuarial information is at odds with the clinical judgment. The results indicated that practitioners are more likely to override actuarial information that indicates a low risk of harm rather than a high one, confirming the existence of the 'precautionary principle'. The research also produced some important messages for practice, particularly a reluctance to reduce sexual offenders' risk of harm even when evidence of all types was compelling, and conversely, a willingness to reduce non-sexual offenders' risk on the basis of only flimsy dynamic evidence, and counter to actuarial pointers. The research concludes that a more sophisticated understanding of the evidence around dynamic factors would enhance assessments.

**Keywords:** risk; assessment; actuarial; clinical; static; dynamic

Probation practitioners are not unique in being expected to predict the future. Social workers predict whether children will be safe living with their families, medics predict how symptoms will develop, and workers in the financial sector predict movement within markets. All these tasks have inbuilt uncertainty, and workers within all of those fields will be versed in the relative merit of prediction based on standardised, statistically-based tools, and prediction based on professional judgment.

The financial sector provides a contemporary cautionary tale about the consequences of placing too much faith in a statistical model of prediction; tools such as the Value at-Risk (VAR) failed to predict the recent sudden downward trajectory of the markets. Indeed the book *Fooled by Randomness* by Taleb (2004) has argued that our universal tendency to perceive the world as predictable obscures an unsettling reality about the way the financial markets behave – success on the trading floors (and indeed life) is largely down to luck. Mervyn King, Governor of the Bank of England seemed to concur when he recently commented thus on the economy:

Anyone who thinks they can tell the future is deluded. (King 2009)

This is a counsel of despair not shared by the probation service. A more meaningful disclaimer for them might be:

Anyone who thinks we can always tell the future is deluded.

For the best part of two decades, the notions of ‘risk assessment’, ‘risk management’ and ‘public protection’ have come to represent the key aims of the probation service. The task of judging those offenders who pose a real and present risk of harm (RoH) to others, and then managing that risk, has come to dominate. Moreover, that judgment governs the intensity with which an offender is worked, and decides whether an offender will be worked with under the multi-agency public protection arrangements (MAPPA). The evidence could be stacked up to show probation practitioners are really rather successful in this task. For instance, my own research (Ansbro 2006) found that although high-risk offenders occupied only 4% of the caseload at the time, offenders in that risk category produced a full quarter of a sample of serious further offences (SFOs). National probation service research (National Probation Service 2005) found a similar pattern, but as an example of the ‘cup half empty, cup half full’ principle, this is rarely reported as a ringing triumph for accuracy in risk prediction. Indeed, it is more likely to be seen as a confirmation of the propensity for SFOs to emanate from all risk bands, meaning variously that risk is dynamic, or that risk assessments lack accuracy, depending on the argument being made.

Given that the task has become central, the approaches used by probation staff are sensibly eclectic. The probation service has always advocated a judicious combination of actuarial and clinical approaches, and, as already stated, both approaches have their strengths and limitations. An actuarial tool will be transparent, consistent, evidence based and relatively unbiased. Moreover, when put to the test, actuarial prediction is slightly more accurate than clinical judgment, a feature that will be returned to shortly. However, it will predict the most likely behaviour of a group, and will not be sensitive to the infinite idiosyncrasies of the individual. It will be impressively accurate at predicting commonplace events (be it run-of-the-mill burglaries or a simple streptococcal infection) but will not pick up on rare but grave situations (be it a predatory paedophile about to offend, or a case of necrotising fasciitis about to break out). Clinical judgment will, at its best, use knowledge, wisdom and skill to tune in to the nuances of the individual and the situation, but at its worst will be biased and based on the assessor’s hunches and prejudices rather than evidence.

## Defining the Terms

Before proceeding, the subtle but significant distinctions in the use of language in this area need to be sorted out. Clinical and actuarial approaches to risk assessment are sometimes seen as clinical judgment is based solely on social and psychological factors that can change. It is true that the universal actuarial predictor used in the probation service, the offender group reconviction scale (OGRS) uses only historical, static information. One of its limitations is that it synonymous with dynamic and static factors, as if actuarial tools just use static information, and does not include dynamic factors, and, therefore, cannot reflect improvements or deteriorations in relationships, drug habits, employment, attitudes and so on. However, this is not so for many other actuarial tools. The historical clinical risk tool (HCR20) for instance is actuarial yet draws on both static and dynamic factors (for example, levels of impulsivity and insight as well as history of offending and mental illness – the former might change over time, the latter cannot). Similarly, Sections 3–12 of the offender assessment system (OASys) guide the interviewer through a series of dynamic factors – social and psychological factors that can change (Home Office 2007). These are translated into actuarial scores, weighted so as the criminogenic factors with the strongest statistical relationship with reoffending (for example, thinking and behaviour) carry a higher score than others (for example, accommodation). Equally, a clinical assessment might draw on statistical knowledge as well as dynamic. A practitioner could, for instance, be assessing a man to be sentenced for downloading internet pornography, and introduce into their thinking some research-derived statistics on the proportion of offenders with that conviction who go on to commit more contact sexual offences (such as Seto and Eke in 2005, who said around 4%). As Parsloe (1999) pointed out, it is somewhat artificial to see actuarial and clinical information as mutually exclusive.

### Actuarial and Clinical Assessment: Which is More Accurate?

Because this research attempted to probe the balancing of actuarial and clinical, the comparative predictive power of the two approaches merits further elaboration. The superior predictive power of actuarial prediction over clinical judgment is a finding that is replicated across a great deal more than the criminal justice literature. A meta analysis by Grove *et al.* (2000) looked at 136 separate pieces of research, comparing actuarial indications with professional/clinical judgment in diverse settings, including the prediction of numerous medical and psychiatric conditions, recidivism and performance on probation, job turnover, the effect on sales of magazine advertising, lie detection and detection of malingering. Consistently the actuarial models achieved greater accuracy, although admittedly there was a small scattering of instances where the reverse was true. Generally the message was clear; actuarial out-performs clinical. However the caveats to this are: (i) the difference in accuracy between actuarial and clinical methods ranges from a whisker to a substantial difference; and (ii) the accuracy of actuarial prediction is still rarely anything to write home about.

The criminal justice literature has had to take a side-step into allied disciplines to measure in more detail this gap in accuracy between actuarial and clinical approaches, as none

has been undertaken using probation staff as subjects. OGRS as a tool in its own right has been evaluated and found to compare well to other tools; see, for instance, Maden *et al.*'s (2006) research. However, it has become unfashionable to evaluate the professional judgment of probation staff; one could speculate that before the 1990s probation staff were not expected to predict anything, and once the 1990s arrived it was expedient to conclude that their clinical judgment was poor. The clinical judgment of psychiatric staff, however, has been put to the test. Despite the differences in training and role between probation and psychiatric staff, the similarities (a grounding in the social sciences, working with people to achieve change) are sufficient to make such a comparison meaningful. Here are two well-cited pieces of research in the mental health field that are probably characteristic of the gap in accuracy between clinical and actuarial prediction.

Lidz, Mulvey and Gardner (1993) evaluated psychiatric staff's clinical prediction of violence in a mentally ill population. He followed up a group of 357 psychiatric patients who had been assessed by the psychiatric team in the emergency room as posing a high RoH following discharge. He had an equal-sized control group, matched for age, sex, race and reason for their admission, but who had not been identified by the same clinicians as posing a high risk. Six months later, 53% of the assessed group had behaved violently, compared to 36% of the control group. Although that represents a substantial gap, 47% of the group were falsely predicted to behave violently ('false positives'), leading Lidz and his colleagues to conclude: 'clinicians are relatively inaccurate predictors of violence to others' (p.269). Turning now to put actuarial prediction on the stand, Monahan (2000) appraised the violence risk appraisal guide (viewed as one of the most accurate actuarial predictors of violent recidivism) by labelling its higher categories as 'high RoH' and its lower categories as 'low RoH'. After a follow-up of three years he found that 55% of the high-risk offenders had reoffended violently, compared to only 19% of the low-risk. This is an impressive success rate, with a bigger gap than Grove's clinicians managed. However, 45% of the high-risk group did not offend violently (false positives), defying prediction. Translating this into guidance for probation practitioners, we can summarise by saying that actuarial prediction beats clinical – but by a modest amount. It is the subtlety of that message that can be difficult to convey.

### **Static and Dynamic Information: Which is Most Important?**

Shifting the focus on to the amalgamation of static and dynamic factors, practitioners find themselves required to get it right, but lacking a precise formula by which to do so. In 2005, HM Inspectorate of Probation (HMIP) carried out a thematic inspection of OASys, when they concluded by offering this formula:

HMI Probation has found evidence from this inspection and from ESIs that the continued over-reliance on the static (actuarial) OGRS2 as an indication of risk of reconviction has meant that the dynamic nature of the OASys likelihood of reoffending assessment is underplayed in managing offenders. (HM Inspectorate of Probation 2005, p.16)

However, as we shall see, over the subsequent few years, and largely based on reviews of SFOs, the message has been exactly the opposite. The precise significance of static actuarial information relative to dynamic comes into focus when we consider the case of Damien Hanson, one of the two co-defendants who murdered John Monckton in November 2004. Hanson was aged 17 years in 1998 when he received a twelve-year custodial sentence for attempted murder and conspiracy to rob. In 2003 he was coming up to his first parole application, and the reports prepared for this were almost unanimous in their support. His home probation officer was impressed by his improved motivation and maturity and assessed him as posing an intermediate RoH. The prison cited his improved behaviour – his pattern of adjudications for assaults earlier in his sentence had given way to co-operation – and a clutch of NVQs and GCSEs. The seconded probation officer cited attitudinal change achieved largely through attending accredited programmes, along with behavioural change around the prison. In summary, the support was based solely on a clinical appraisal of dynamic factors. A full RoH assessment was done days after the parole assessment report, in which an OGRS score of 91% is recorded, but the information never became part of the overall formulation of risk. The report from the independent Parole Board member was the one dissenting voice; parole was not supported because of Hanson’s unrealistic plans for his release, his minimisation of his offending, and the fact that he had not, at that point, completed an anger management course. Parole was refused. A year later, with the anger management course completed and no input from an independent Parole Board member, parole was granted. By now a full OASys had been completed featuring, as before, an OGRS score of 91%, accompanied by a risk of sexual or violent harm banding of ‘high’. None of this information featured in the second parole application either. What difference would its presence have made? Perhaps Hanson would have been released at that point anyway or perhaps the next year. However, it may very well have ensured that he was referred to MAPPA before his release, and managed with a firmer grip afterwards.

Other aspects of the management of this case have been identified and criticised elsewhere (the licence conditions, the appropriateness of a particular anger management course) but the important point of relevance as far as this research is concerned was that actuarial information was absent from the decision. It was a failing that ultimately led to Hanson being supervised by a team who worked with high volumes of lesser-risk offenders than would have been the case had he been released with a ‘high risk’ label. The point was given a good deal of attention by the Inspectorate; moreover they cite a Parole Board newsletter from the year before the SFO which echoes the same sentiment, based on a review of parolees who were recalled:

Members are not giving enough weight to the *static* risk factors before balancing them with the *dynamic* risk factors. (Parole Board 2005, cited in HM Inspectorate of Probation 2006a, p.30, italics in original)

A change in formula from the HMIP’s advice from the previous year, but the case gives us a digestible, straightforward message – ensure that sufficient notice is taken of static, actuarial information, and do not attach an exaggerated importance to dynamic factors. Shortly after the

Hanson and White case, HMIP had cause to publish a review of another SFO, on Anthony Rice, a parolee who murdered Naomi Bryant in 2006. In it a similar message was drawn out:

The risk faced by all staff involved in a case such as this is that they can sometimes overvalue the progress made on the *dynamic factors* in relation the *static* ones. An able offender manager should be able to retain the necessary long view, keeping the *static* and *dynamic* factors in proper proportion with each other. (HM Inspectorate of Probation 2006b, p.56, italics in original)

However, the 'necessary long view' in this case would require psychic powers. HMIP's criticism surrounds the decision in 2001 to allow Rice to move from a Category B prison (Grendon) to a Category D prison (Leyhill, an open prison). This was a critical point in his progress through the life sentence, signifying movement towards eventual release. At this stage the 'Structured Risk Assessment' (SRA-2000), an actuarial tool based on static historical factors, was administered by the prison psychologist. It indicated a 'very high' likelihood of reoffending – more specifically a 72% chance of sexual reconviction within 20 years. Waiting for this score to come down was futile – the whole point about static factors is that they cannot change. The dynamic factors, however, had changed, and were fairly favourable; in the eleven years since starting his sentence he had followed various treatment programmes, and had:

shown evidence of taking responsibility for his behaviour, of increasing his level of victim empathy and of altering his distorted attitudes towards women. (HM Inspectorate of Probation 2006b, p.24)

Over recent years our understanding of sexual offenders' dynamic risk variables has expanded, evolving mainly from Hanson's and Harris's (2001) model that defines two types of dynamic factors: stable and acute. However, progress cannot be demonstrated on some of these areas whilst the offender is in custody, for instance, intimacy deficits and social contacts. At some stage something of an act of faith is required, which goes something like this: does the offender's change in attitude indicate a profound shift that will genuinely decrease the chance that they will do it again, or are they just pretending? HMIP's criticism could be accused of displaying a hindsight bias.

Neither does the recent case of Dano Sonnex offer resolution to these complexities (Ministry of Justice/Hill 2009). He was an abused child, excluded at school, and a heavy drinker and crack user by his teens. However, he had no criminal convictions until, at the age of 17 years he was sentenced to eight years in custody for robbery and wounding, using an imitation gun and a knife. In youth custody, Sonnex was regularly violent to other inmates and prison staff, set fire to his cell on one occasion, and showed signs of paranoia on the wing. Towards the end of his sentence he had calmed down, but he was still refused parole twice. Throughout this, his risk assessments are confused. The OGRS score predicted just a 33% chance of reoffending within two years (he had no previous convictions and OGRS works on a number of static factors predominantly around previous offending), the e-OASys put him at high risk to the public and a known adult, whereas the record of contact (known as Delius) on his release from custody

pegged him as a medium risk, and tier 3. The assessment of risk would not have affected whether or not he was released at that point, as his two applications for parole were refused, and he was released at the latest possible date. However, if Sonnex had been clearly labelled a high-risk, tier 4 offender he would have been managed by a specialist team, with the resources to plan for his release with care, and able to ensure the involvement of MAPPA. As it was, he was held by a team for medium- and low-risk offenders, where his probation officer's caseload totalled 127. In June 2008, four months after his release, Sonnex brutally murdered two French students. This was a case where a catalogue of minor errors were pieced together after the event; however, for the purposes of this research it shows the impossibility of compiling rules around the integration of static and dynamic factors; in this case the OGRS score was almost irrelevant, and the set of dynamic, criminogenic factors that feature in OASys all important.

### **This Research**

Perhaps the only real conclusion that can be drawn from this is that amalgamating the clinical, the actuarial, the static and the dynamic is a tricky business; how do we put it all together and come up with the best formulation?

This research questioned what it took for practitioners to override the actuarial prediction provided by OGRS, the one actuarial tool that is used by probation practitioners on all offenders. OGRS is a Home Office product, first launched in the early 1990s, and now in its third version. Each version has made refinements on the last, and draws on a larger sample size – 30,000 for OGRS2, 79,000 for OGRS3 (National Offender Management Service, undated guidance on OGRSs). OGRS uses information about the age, gender and offending patterns of the cohort to predict, in percentage terms, the probability that an offender will be reconvicted within one and two years. OGRS3 claims to measure reoffending rather than reconviction because it takes into account warnings and cautions as well as convictions – this seems rather an ambitious claim as it cannot include the more substantial number of offences that are committed and undetected. This research was based on OGRS2, as OGRS3 was just starting to be used when the data were collected. As well as the prediction of reconviction, OGRS2 contained an indicator of the risk of sexual or violent harm. This was a more ambitious project, as these offences occur less frequently and are, therefore, much less predictable. As such, it classified offenders into three broad bands; 'no history' (that is, no previous sexual or violent convictions, therefore very little to go on), 'medium risk' or 'high risk'.

OGRS now forms an integral part of OASys. This is another Home Office project, which is used on all offenders, and is designed to integrate the actuarial with the clinical, the static with the dynamic, and to assess needs as well as risk. It guides the practitioner's interviewing and information gathering through an offence analysis and a range of criminogenic factors to produce a numerical score indicating the likelihood of reconviction (the maximum score is 168). Where indicated, it takes practitioners through a more thorough risk assessment, and alerts workers when a specialist assessment is needed. Practitioners summarise its contents by allocating offenders to one of four RoH categories; 'low', 'medium', 'high' or 'very high'. This

research was interested in those cases where there was a big gap between the OGRS score at the beginning of OASys, and the overall RoH band at the end of OASys.

More specifically there were two questions being asked. The first to be posed was whether practitioners were more likely to override the OGRS assessment of RoH in an upwards or downwards direction. In a climate where underestimating an offender's risk can be dangerous, there is a tendency to consistently err on the side of caution (Kemshall 1998). This precautionary principle would be evidenced if practitioners were more ready to override a low-risk actuarial assessment with a high-risk clinical assessment than override a high-risk actuarial assessment with a low-risk clinical assessment. The second line of enquiry was to identify the reasons in practice for such overriding.

### **Method**

The research was based on the OASys database for a large metropolitan probation area, which, at the time of gathering the data (May 2008), contained 17,927 cases. A small group of atypical cases was isolated: that 2% of the database where there was a considerable discrepancy between the actuarial indication (that is, the OGRS2 prediction of reconviction in percentage terms and its banding of sexual or violent harm), and the eventual RoH assessment that concluded the full OASys assessment. This group as a whole is referred to as the 'discrepancy cases', and can be broken down into two types of case. First there were those cases where the OGRS tool assessed the individual's RoH at the lowest possible level because of an absence of previous sexual or violent offending, and yet the final OASys RoH assessment was of 'high' or 'very high' risk. Second there were those cases that were the other way round, where the OGRS tool pitched the RoH at the highest level, and yet at the end of the full OASys assessment the practitioner put the individual at low RoH. Twenty of each type were picked randomly, and read in detail, in an attempt to establish the reason why the OGRS indication of RoH had not been borne out when all the available information was considered.

I appointed myself as the arbiter of such decision making, and identified cases where the overriding had been done for sound, well-evidenced reasons, where it reflected weak practice, and where a case could be made either way. Before the results are discussed, it must be acknowledged that the judgments made about the risk assessments are qualitative, and, inevitably, have an element of subjectivity to them. This is inevitably the case when evaluating the quality of a risk assessment. There are exceptions, of course; writing unintelligibly, omitting offence details, neglecting previous convictions or failing to do an assessment at all, are all inarguably poor practice. However, beyond those gross failures, classifying assessments as examples of good practice, or of poor judgment necessarily reflects, in part, the researcher's views on what constitutes good practice. There is a paradox that runs through policy and literature on risk assessment. On the one hand there is an acknowledgment that risk assessment is part art, part science, an acceptance that future offending can only be anticipated up to a point, and to some extent defies prediction. This acknowledgement is clear, for instance in statements such as this made by Andrew Bridges, Chief Inspector of Probation in the 2006 report into the murder of Naomi Bryant by Anthony Rice:

... even the most sophisticated assessments can only be an aid to decision-making, not a substitute for it. (HM Inspectorate of Probation 2006b, p.69)

On the other hand there is a theme that identifies some risk formulations as correct, and some incorrect. This can also be found in the tone of HMIP inspections, where a more binary, right or wrong process is implied. In their October 2009 *Risk of Harm Inspection Report: A Stalled Journey*, which surveyed a number of London Boroughs, the Inspectorate concluded:

Inspection staff judged that the *RoH* analysis accurately reflected the risk to children in 85% of cases, to the public in 70% of cases, to a known adult in 80% of cases and to staff in 89% of cases. (HM Inspectorate of Probation 2009, p.13, italics in original)

In the same inspection, practitioners' judgments are deemed to be inaccurate because of factors such as 'insufficient links between problem areas in the offender's life and the *RoH*' (p.13, italics in original) or because the analysis had not made 'appropriate use of previous probation, prison or Youth Offending Team/ Service assessments' (p.14). Such brief statements leave the reader curious about the details, which are never given. If they were given, one suspects that there might be more than one defensible position to take on the subject. Kemshall (2008) sums up this conundrum:

while there are a number of significant issues in 'knowing' high risk or dangerous offenders, policy, legislation and practice are all conducted **as if** we can know them. (p.13)

Therefore, in this research the stance I have taken is this: it is justifiable to take a view on the quality of a risk assessment as long as the reasons for that view are set out.

## Results

Tables 1 and 2 show the results of the study. In these tables:

- High OGRS = 'high risk' of sexual or violent harm (OASys, Section 1.16), and scoring over 80% probability of reconviction within a two-year period (OASys, Section 1.15)
- Low OGRS = 'no history' classification of violent or sexual offending (OASys, Section 1.16), and scoring under 20% probability of reconviction within a two-year period (OASys, Section 1.15)
- High RoH = 'high' or 'very high' overall RoH on OASys (Section R 10.6)
- Low RoH = 'low' overall RoH on OASys (Section R 10.6)

**TABLE 1**

*Breakdown of the OASys Database by Type of Discrepancy*

Total number of OASys in database	High OGRS/low RoH	Low OGRS/high RoH
17,926	91 = 0.5% of OASys database None of these were sexual offenders	271 = 1.5% of OASys database 112 of these were sexual offenders

**TABLE 2**

*Breakdown of the Discrepancy Cases*

Type of case	Good calls	Bad calls	Either way
20 high OGRS/low RoH	9	9	2
20 low OGRS/high RoH	7	8	5

**Findings and Discussion**

*Were practitioners more likely to override the OGRS assessment of RoH in an upwards or downwards direction?*

This question was satisfyingly simple to answer. Out of the 17,926 cases, there were 91 where OGRS indicated a very high risk of sexual or violent harm (see [Table 1](#)), and yet the ultimate risk category was low. This represents 1 in every 197 OASys assessments, or 0.5% of the database. In contrast there were 271 cases where the OGRS indication of sexual or violent harm was low, and yet the eventual risk category was high. This represented 1 in every 66 cases, or 1.5%. The first point to acknowledge is that cases with such a discrepancy between actuarial and clinical information are rare. However, where they do occur, it is quite clear from this sample that practitioners are three times as likely to override actuarial information when it points to low risk rather than high risk. Kemshall wrote about this tendency in 1998, and her conclusions are as apposite now as then: in a culture where that task is to ‘predict the unpredictable’, the precautionary principle dominates practice, ‘operating on a ‘what if’ basis of the worst case scenario, often in the face of low evidence of potential hazard’ (Kemshall 1998, p.68).

*What were the reasons in practice for overriding the OGRS assessment of RoH?*

The second question that the research aimed to answer was why such decisions were being made. Before addressing this, it is important to acknowledge that this research was based on OASys assessments alone, and previous research has suggested that practitioners often know a great deal more about their cases than are contained within OASys (Crawford 2007). Given that caveat, I will set out the main reasons by referring to cases where the decision making appeared to be sound and well evidenced (the 'good calls'), and those where it was not (the 'bad calls'). There was also a small number where a convincing case could have been made either way.

#### Formulations for the high OGRS/low RoH cases

Out of the 20 cases that were examined, nine were judged to be good calls and nine were bad calls (see [Table 2](#)). Two of the good calls were the result of an inflated OGRS prediction caused by a glitch in the actuarial process. More specifically, they were women who were sex workers and had accumulated many prostitution-related convictions. They had been entered on OGRS as sexual offences, which had instantly escalated their RoH; the practitioners had very sensibly de-escalated them. The other cases were, perhaps, more useful in providing lessons for good practice. They were cases where the practitioner had decided that sufficient time had elapsed, or that progress had been made across a sufficiently convincing range of areas to back up the opinion that their RoH had diminished. For example, Andy was aged 58 years and had been involved in organised crime as a young man. His first custodial sentence was at the age of 15 years, and he amassed eleven prison sentences for, variously, burglary, violence and fraud. Eventually he was sentenced to life at the age of 27 years for murdering a police informer. He and his co-defendant were drunk when they murdered their victim, and the attack was brutal, stabbing him in the throat and anus. He served a life sentence of nearly 20 years before release on licence, at which point he had an OGRS score of 80%. At the most recent OASys review, Andy had been out for more than ten years, had stable accommodation, and regular work as an extra on film sets. Moreover he had always co-operated with supervision by good attendance, a willingness to examine and discuss his criminal past and his life post-release. He described occasionally drinking alcohol (and there was no evidence to contradict this) and in all his contact with the various agencies he demonstrated an ability to manage frustrations that presented themselves (for example, in negotiations with the housing department). What was reassuring about this assessment of 'low' RoH in the face of a high actuarial prediction was not only the time that had elapsed, but the range of criminogenic factors that had changed for the better; the decision to pitch his RoH as 'low' was presented as considered and defensible.

It was an absence of this range that was noticeable in the cases where the decision was judged to be faulty. The case of Charlie demonstrates this point. Charlie's OASys described a 43-year-old man who had numerous convictions including violence, burglary and drug-related offences. Hence, the OGRS produced a score indicating a 93% chance of being reconvicted within two years, and placed him in the category posing a high risk of sexual or violent harm. The practitioner's decision to override this and place him at 'low' RoH seemed to stem from two factors. The first was that the offence for which he was on the current community order was relatively trivial – shoplifting to fund his ongoing drug habit. The wisdom of this is doubtful.

Research such as Craissati's and Sindall's (2009) recent review of SFOs observes the importance of keeping in mind the whole picture, and not relegating previous background noise of previous convictions of violence. The second factor was that he had recently started attending a community treatment facility for his drug misuse. The practitioner's willingness to interpret involvement in treatment as an indication of meaningful reduction in risk is well-documented. For instance, in the Inspectorate's report into the management of Damien Hanson (HMIP 2006a) an over-reliance on dynamic factors and a relative disregard of static factors is criticised. It is easy to sympathise with the practitioner's dilemma; on the one hand the requirement to refer to accredited programmes and partnership agencies is based on a message that their efficacy is evidence based. Practitioners are then criticised if they then assume that completion of such programmes has achieved significant change. What is perhaps missing is an appreciation of the scale of change achieved by accredited programmes, and the complexities that abound. A simple statement about interventions that 'work' fails to appreciate that a good treatment effect is likely to be a reduction overall of around ten percentage points and that findings of effectiveness are not always replicated. Returning to Charlie's case, it is possible that the practitioner had sensed the beginning of a true desistance curve in this man, but the 'low' RoH assessment because of his trivial index offence and early signs of progress on a narrow range of criminogenic factors seemed over-optimistic.

#### *Practitioners' formulations for the low OGRS/high RoH cases*

As stated, the cases where actuarial guidance indicating 'low' RoH was overridden in favour of a 'high' RoH were more frequent, an unsurprising trend given the anxiety practitioners experience. Out of the 20 cases that were examined in this category there were seven that seemed to be well thought out and thoroughly backed up ('good calls', eight that were not ('bad calls') and five that could be argued either way (see [Table 2](#)). Taking the good calls first, there were two reasons that appeared to be guiding practitioners' thought processes. First, there were cases where there had been careful intelligence gathering, and second, there were cases where the index offence was characterised by revenge, or a clear psychological vacuum around the circumstances of the offence.

The first of these themes is clear in the case of Omar, a 52-year-old man who was on a suspended sentence order with a condition to attend the sex offender treatment programme. He had been found guilty of touching a ten-year-old girl's breasts while at a market. His defence was that he had been wrongly identified as the culprit. OGRS gave him a 4% probability of reconviction, and categorised him as the lowest risk of sexual or violent harm because he had no history of such before the index offence. However, the probation officer had spoken to a colleague in the Child Abuse Investigation Team, and discovered a considerable record of alleged sexual offences against children that had not resulted in conviction, including three sexual assaults against girls, and the buggery of a four-year-old boy. Omar had found most of his alleged victims through his job as an airport cleaner, allowing him access to children alone in the airport toilets. Moreover, these children were about to leave the country and were unlikely to

come back to give evidence against him. Although the use of intelligence in risk assessment and MAPPA is something that can be troubling, in this case surely the most vehement defender of due process would concede that it was justifiable and invaluable. It resulted in a 'high' RoH, and provided material that helped the treatment programme make real progress with Omar.

The other theme that characterised the 'good calls' was offending motivated by jealousy, along with a lack of affect and cognition around the offence. This is illustrated by the case of Alan, whose OGRS probability of reconviction was 14% and his risk of sexual or violent reconviction was in the lowest category. He had set fire to his girlfriend's flat when she was attempting to end their relationship, in favour of a new partner. Alan could not access any memory of these events, or perhaps was not prepared to acknowledge their existence. Either way, the probation officer was not able to examine his feelings about his behaviour, the extent of his planning, or his ability to appreciate its consequences. Prins (1999) succinctly surveyed the literature on this subject, and he concludes that the capacity to act on feelings of possessiveness or rejection should always be taken seriously. He also reminds us of his list of 'seven sins of omission' (originally set out in Prins (1995), but elaborated on in later publications, for example, Prins (2005)). One of his 'sins' is the practitioner's failure to examine whether the offender has come to terms with what they did. In Prins's view this has to be more than merely expressing remorse, which is easy to fake and can be relatively meaningless. Even worse, it may inspire the offender to find a way of offending which deflects blame onto the victim or a third party. The offender must make progress in contemplating and reflecting on their offending, a more demanding and discursive process than merely expressing contrition. Alan could not or would not make a start on this process, and the practitioner's decision to place Alan in the 'high' RoH category seems well-justified.

Turning now to the eight cases judged to be 'bad calls', they were largely examples of the precautionary principle – a questionable inflation of risk. The phenomenon seemed particularly intractable with sexual offenders. Tony was a 30-year-old man on a three-year community order with a condition to attend the sexual offenders' treatment programme, after pleading guilty to downloading child pornography. His OGRS probability of reoffending was 13%, and his risk of sexual or violent offending was in the lowest category. The practitioner had, however, placed him in the 'high' risk category at the start of his order – understandably – but had, more surprisingly, resisted lowering his risk categorisation for the duration of the order. This was despite Tony's progress in a variety of important areas. The past decade has seen an abundance of research into proven dynamic risk factors that extends our understanding way beyond the early preoccupation with static factors or pro-offending attitudes. Hanson and Harris (2001) were the initial protagonists in this area, but variations on their model have since followed (for instance, Beech *et al.* 2002; Thornton 2002). Craissati and Beech (2003) reviewed the literature on this subject, and synthesised the various evidence and models into five stable dynamic areas and four acute factors that are significant in reducing risk. The stable dynamic factors are intimacy deficits, social influences, pro-offending attitudes, self-regulation of a general and sexual kind, and deviant sexual interests, and the four acute factors that can indicate a more sudden escalation of risk are substance misuse, negative mood, anger/hostility, and opportunities for victim access. Over recent years the Ministry of Justice has developed this

into a formalised risk assessment tool for workers in the criminal justice system, using the same categories, separated in the same way into stable and acute factors, with an additional emphasis on response to supervision. Tony had made progress on all of the stable factors, with no evidence of the acute 'crisis' factors. He had finished the treatment programme, impressing the leaders with his commitment and frankness. As suggested earlier, it would have been unwise to set too much store by this on its own, but in Tony's case he had managed disclosure to one employer (he was a lab technician), cope with losing that job, and through voluntary work had gained another job with a charity (again, with appropriate disclosure). He had improved his social functioning by spending less time on his own, becoming involved in social functions through work, and discussed his struggle to reduce fantasies of children of his preferred age.

Tony's case brings us to an important final point. In this research there were no sexual offenders in the 91 cases where a high actuarial score was overturned in favour of a 'low' RoH. In contrast there were 112 sexual offenders in the 271 cases where a low actuarial score was overturned in favour of a 'high' RoH. This research considered a small sample of unusual cases, but it provides a snapshot of trends in risk assessment that we could assume to be more prevalent in the rest of the OASys database. If risk is truly dynamic, then it must be for sexual offenders as it is for other offenders. The probation service cannot afford to trivialise the public's concern about sexual offending, or underestimate the risk of recidivism; however, neither should it allow their risk to be escalated, or only take notice of the extensive evidence base when it indicates that risk is high. This pattern has characteristics of 'penal marking' (Hudson 2002), the phenomenon that brands sexual offenders as a group for which no measures are too extreme, with no rights deserving to be upheld. Sexual offenders are the most researched and written-about group of all offenders. This wealth of knowledge may be being disregarded in practice, in favour of a heuristic that clubs all sexual offenders together, labelling them resistant to change, and more likely to reoffend than the evidence indicates.

### **Conclusion**

The results of this research could be condensed into three key areas. First there is the finding that practitioners were more inclined to override a low actuarial score than a high one. This confirmation of the 'precautionary principle' in action is entirely unsurprising. As Kemshall (1998) identified over ten years ago, practitioners face the mutually-exclusive targets of high accuracy and high throughput, and exist in a climate where failings in practice will be hunted for if an offender commits a serious offence whilst on supervision. In this invidious situation, it is to be expected that practitioners err on the side of caution, and find it easier to override a low OGRS score than override a high one.

Second, there is the more detailed discussion of factors that guide their decision making. There was evidence of much impressive and skilful practice in the assessments that were read, but weak formulations of risk were commonly those that were based on too narrow a range of dynamic factors.

The third finding is linked to the last point: there was a striking contrast between assessments for sexual offenders and non-sexual offenders. Even when the OGRS score was low, and was accompanied by evidence from a range of dynamic factors, practitioners seemed to be reluctant to reduce sexual offenders' risk. In contrast, there was evidence of some over-eagerness to override a high OGRS score where non-sexual offenders were concerned, resulting in an offender's risk being downgraded without a convincing range of dynamic factors to support such a decision. Balancing the static and dynamic in the risk assessment process is a theme that has cropped up consistently in the literature on SFOs; the message here is for practitioners to consider the full range of risk factors that OASys offers; completing accredited programmes on its own is not sufficient.

Practitioners are subject to contradictory, opposing pressures. First, research on effective practice confirms that certain dynamic factors are criminogenic, and if they change for the better, then chances of reoffending and inflicting harm are reduced accordingly. However, practitioners are vulnerable to criticism when they apply this rule too simplistically in their risk assessments. On the other hand the literature, understandably, urges them to resist 'the precautionary principle' to avoid universal risk inflation, and the clogging up of MAPPA and prisons. The partial, and not quite satisfactory, solution that this research suggests is that decisions should never be made on a single dynamic factor; if a sexual offender has achieved change on the full range of criminogenic factors then the decision to lower risk is defensible and should be made confidently. Conversely, if a non-sexual offender has made progress on only one or two factors, then exercise caution.

### References

- Ansbro, M. (2006) 'What can we learn from serious incident reports?', *Probation Journal*, 53(1), 57–70.
- Beech, A., Friendship, C., Erikson, M. and Hanson, R.K. (2002) 'The relationship between static and dynamic risk factors and reconviction in a sample of UK child abusers', *Sexual Abuse: A Journal of Research and Treatment*, 14(2), 155–7.
- Craissati, J. and Beech, A. (2003) 'A review of dynamic variables and their relationship to reconviction', *Journal of Sexual Aggression*, 9(1), 41–55.
- Craissati, J. and Sindall, O. (2009) 'Serious further offences: an exploration of risk and typologies', *Probation Journal*, 56(1), 9–27.
- Crawford, A. (2007) 'What impacts on quality assessment using OASys?', *Probation Journal*, 54(2), 157–70.
- Grove, W.M., Zald, D.H., Hallberg, A.M., Lebow, B., Snitz, E. and Nelson, C. (2000) 'Clinical versus mechanical prediction: a meta-analysis', *Psychological Assessment*, 12(1), 19–30.

- Hanson, R.K. and Harris, A.J.R. (2001) 'A structured approach to evaluating change among sexual offenders', *Sexual Abuse: A Journal of Research and Treatment*, 13(2), 105–22.
- HM Inspectorate of Probation (2005) *Realising the Potential: A Short Focused Inspection on the Offender Assessment System (OASys)*, London: HM Inspectorate of Probation.
- HM Inspectorate of Probation (2006a) *An Independent Review of a Serious Further Offence Case: Damien Hanson and Elliot White*, London: HM Inspectorate of Probation.
- HM Inspectorate of Probation (2006b) *An Independent Review of a Serious Further Offence Case: Anthony Rice*, London: HM Inspectorate of Probation.
- HM Inspectorate of Probation (2009) *Risk of Harm Inspection Report: A Stalled Journey*, London: HM Inspectorate of Probation.
- Home Office (2007) *OASys Handbook*, London: Home Office.
- Hudson, B. (2002) 'Human rights, public safety and the probation service: defending justice in the risk society', *Howard Journal*, 40. 103–13.
- Kemshall, H. (1998) 'Defensible decisions for risk: or "it's the doers wot get the blame"', *Probation Journal*, 45(2), 67–72.
- Kemshall, H. (2008) *Understanding the Community Management of High Risk Offenders*, Maidenhead: Open University Press.
- King, M. (2009) Press interview on the UK quarterly inflation figures, 13 May.
- Lidz, C., Mulvey, E. and Gardner, W. (1993) 'The accuracy of predictions of violence to others', *Journal of the American Medical Association*, 269, 1007–11.
- Maden, A., Rogers, P., Watt, A., Lewis, G., Amos, T., Gournay, K. and Skapinakis, P. (2006) *Assessing the Utility of the Offenders Group Reconviction Scale-2 in Predicting the Risk of Reconviction within 2 and 4 Years of Discharge in English and Welsh Medium and Secure Units*, National Forensic Mental Health Research and Development Programme report. Liverpool: University of Liverpool.

- Ministry of Justice/Hill, L. (2009) Investigation into the Issues Arising from the Serious Further Offence Review: Dano Sonnex, London: Home Office.
- Monahan, J. (2000) 'Violence risk assessment: scientific validity and evidentiary admissibility', Washington and Lee Law Review, 57, 901–18.
- National Probation Service (2005) Learning Points from Serious Further Offence Full Reviews, London: National Offender Management Service/National Probation Service.
- Seto, M.C. and Eke, A.W. (2005) 'The future offending of child pornography offenders', Sexual Abuse: A Journal of Research and Treatment, 17, 201–10.
- Parsloe, P. (Ed.) (1999) Risk Assessment in Social Care and Social Work, London: Jessica Kingsley.
- Prins, H. (1995) 'Risk assessment: seven sins of omission', Probation Journal, 42(4), 199–201.
- Prins, H. (1999) Will They Do It Again? Risk Assessment and Management in Criminal Justice and Psychiatry, London: Routledge.
- Prins, H. (2005) Offenders, Deviants or Patients?, 3rd edn, London: Taylor and Francis.
- Taleb, N.N. (2004) Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets, London: Thomson.
- Thornton, D. (2002) 'Constructing and testing a framework for dynamic risk assessment', Sexual Abuse: A Journal of Research and Treatment, 14(2), 139–53.

Date submitted: December 2009

Date accepted: January 2010