Violent Offenders. Appraising and Managing Risk

By Vernon Quinsey, Grant Harris, Marnie Rice, and Catherine Cormier. 2nd edition. Washington, DC: American Psychological Association, 2006. 462 pp. $49.95.

Treating Violence: A Guide to Risk Management in Mental Health

By Anthony Maden. Oxford, UK: Oxford University Press, 2007. 189 pp. $55.00.

Violence Risk. Assessment and Management

By Christopher Webster and Stephen Hucker. Chichester, UK: John Wiley and Sons, 2007. 202 pp. $45.00.

Reviewed by Alec Buchanan, PhD, MD

In *The Hedgehog and the Fox*, Isaiah Berlin[1] developed a metaphor from a fragment of poetry by the Greek poet, Archilochus: "The fox knows many things, but the hedgehog knows one big thing." Berlin's metaphor concerned how people lived their lives. Plato, Dante, Pascal, Nietzsche, and Proust, he thought, had been hedgehogs, their worlds built around a single organizing principle. Aristotle, Shakespeare, and Joyce, on the other hand, had been foxes, eclectic in their thinking and requiring no "unitary inner vision" (Ref. 1, p 2). Berlin's metaphor reappears here courtesy of Webster and Hucker, but with a new moral twist. Hedgehogs continue to know one big thing, but this is now a problem for them—and for other people. They extend the explanatory reach of their one big thing aggressively into new domains, they are impatient with those who do not "get it," and they try to mow their opponents down. Foxes, on the other hand, are flexible and modest and make more accurate predictions. Power has moved to the fox.

For psychiatrists in clinical practice, the belief that the best risk assessments are rooted in the skills they learned as trainees is a bit like the hedgehog's one big thing. This belief is linked, I think, to another widely held and deep one: that understanding the patient is a necessary prerequisite to looking after that patient properly. This "clinical" method has been seen by doctors as, among other things, the best means of assessing and managing the risks that accompany treatment, risks that include the patients' harming themselves or, less commonly, others. There are few signs announcing change of this ethos. The most recent editions of the major psychiatric texts include lists of factors to be borne in mind when assessing risk, but beyond this prescribe little by way of structure to guide the assessment. The textbooks emphasize instead the ways in which good clinical technique links risk assessment and clinical management. Understanding the patient not only helps clinicians assess risk, but also helps them provide better care.

The three books reviewed herein challenge this approach, offering as alternatives several varieties of structured risk assessment. All of these structured assessments contain lists of items to be scored by an assessor. What happens next depends on which approach one chooses. Actuarial assessments combine the scores on each item into a total that is intended to reflect the risk. Structured professional judgment (SPJ) approaches, in addition to including items designed to aid clinical management, give the assessor freedom to include additional information before reaching a conclusion. Different instruments are recommended for different types of risk, including sexual violence and other types of assault, for different age groups, particularly adolescents, and according to the relationship of the potential victim to the client. The dozens of different methods resemble the many small things. If power has indeed moved to the fox, clinicians have some serious thinking to do.

All three books, perhaps predictably, continue an argument among advocates of the different structured approaches, and among the advocates of structured and unstructured approaches, that has been colorful at times. Quinsey *et al.* created the VRAG

(Violence Risk Appraisal Guide), one of the most accurate actuarial instruments, and *Violent Offenders* is uncompromising in its advocacy of actuarial methods. Clinical judgment, at least as it applies to risk assessment, is portrayed as irredeemably flawed by the cognitive biases to which we are all prone. The advocates of SPJ are mocked, for some reason in cod French (they don't know how to "meau de lawn"), and those who support the clinical method are denying evolution. Ouch. Webster is one of the authors of the HCR (Historical, Clinical, and Risk Management)-20, perhaps the best known example of an SPJ approach, and might be expected to defend himself, but *Violence Risk: Assessment and Management* has an informal, sometimes chatty, style, with three-page chapters and many clinical vignettes, and is unsuited to this kind of debate. Not to fear. Anthony Maden calls the VRAG an "elderly, lovable but dementing aunt." Ouch, backatcha, as an Alaskan governor might say.

But alongside the badinage there are surprises for those who see the distinction, between those who favor actuarial approaches, on the one hand, and clinicians, on the other, as black and white. All three books draw on extensive clinical experience. The actuarial orientation of *Violent Offenders* does not prevent it from containing what is, for my money, one of the most clinically resonant contributions, a detailed account of the emotional and relationship instabilities affecting recidivist offenders in the period immediately preceding their crimes. And *Treating Violence*, which sees parallels between the preventive detention of "psychopaths" and restrictions to limit the spread of infectious disease, would nevertheless allow the same psychopaths to choose prison over hospital and hence to receive, in many cases, a desert-based sentence that returned them to the community in relatively short order. Maden's willingness to consider unexpected alternatives should give us all pause. Risk assessment and management require the balancing of many legitimate principles. If the answers to the questions raised by these books were obvious, we would presumably know them by now.

The analogy between the detention of "psychopaths" and infectious disease control is intriguing, but I am not sure that it comes out the way Maden implies. When 4,000 cases of typhoid were being reported in New York every year and there was no effective treatment, three asymptomatic carriers were detained. One of these was Mary Mallon ("Typhoid Mary"), who lived in a cottage on the grounds of an isolation hospital on the East River, and the other two were at home.[2] These arrangements are very different in scale and quality from those that were initially proposed for "dangerous psychopaths" in the United Kingdom. The difference is particularly striking in light of the level of harm caused: at least three people were known at the time to have died after Mallon prepared food for them, two after she had broken the conditions of an earlier release by returning to her trade (she was a cook) under a false name. Other diseases were more actively quarantined before treatment was available, but the risks were often colossal. One estimate is that smallpox infected half of those who shared a house with the victim and that one third of those infected died.[3]

Structured Instruments and Risk Assessment

The structured instruments reviewed here have some common features. First, each provides a list of items and, therefore, can act as an *aide mémoire* to the forgetful clinician. It seems a laudable aim, if a somewhat limited one, given the time and effort that has been expended. Second, they organize information. Reviewing the PCL-R (Psychopathy Checklist: Screening Version-Revised) and its 20-item, 0-1-2, scoring framework that was to be copied by the HCR-20, Webster and Hucker argue that these instruments are of greater value than simple checklists because the items are "definable" and come at you in "manageable" numbers. As opposed to the potentially infinite number of hard-to-define clinical considerations that risk assessors otherwise have to bear in mind. Again, it seems laudable, although I suspect that we do not yet have the data to show that clinicians dealing with well-defined considerations manage risk better than clinicians using hard-to-define ones. The various approaches described in these books also share a large number of items. The HCR-20 includes as variables, "early maladjustment," personality disorder, PCL-SV (PCL: Screening Version) score, substance use problems, relationship instability, and prior supervisions failure. The VRAG includes "problems at junior school," personality disorder, PCL-R score, alcohol abuse, separation from parents before 16, and failure on prior conditional release.

So how do they differ? The first clue to understanding the debate between the advocates of the various approaches lies in their names. As the phrase "structured professional judgment" implies, methods such as the HCR-20 leave a lot of discretion, in assessing and managing risk, with the clinician. The items are to be scored in each case, but not necessarily added up. Having scored each of the 20 items, the clinician makes a summary judgment, allocating the client to one of three risk categories. This process differs, as I understand it, from simple addition in two ways. First the clinician is entitled to attribute extra weight to a single item if that item seems particularly important in an individual case. So, if a man with schizophrenia has previously, and only, been violent when his delusions are present, and his delusions are present now, a positive score on active symptoms of mental illness (C3) can put him in the high risk category even if his overall score is low. Second, the clinician can use information not covered by the HCR-20. So if the same man had bought a gun, that could have the same effect.

This discretion, to use information differently in different cases and to use data not covered by the instrument, is viewed less sympathetically by the advocates of actuarial approaches. Their case seems to be, if there is no empirical reason to add an item, there is no reason at all. I have some sympathy. The VRAG has long raised eyebrows for its inclusion of mental illness as a protective factor. There are no similar surprises in the HCR-20. But is this a good thing? Webster and Hucker tested the HCR-20 on patients in the MacArthur risk assessment study. While some items, including previous violence (H1), Hare psychopathy (H7), and personality disorder (H9) correlated strongly with violence after discharge, others, such as relational instability (H3) and employment problems (H4), showed no association. One HCR-20 item, major mental illness (H6), was protective. Game to VRAG? The finding seems likely to be real: major mental illness is here being compared with personality disorder and substance abuse, and my reading of the literature is that, in such a comparison, patients in the major mental illness category may, indeed, have lower offending rates on discharge, especially if they are being treated.

A second concern of the actuarialists has been, where would this granting of discretion end? Bring in enough discretion and pretty soon you no longer have a structured approach. But it does seem to fly in the face of common sense that a man who is shouting threats to kill his wife after discharge from a hospital should be treated in the same way as someone who is not, just because their scores are identical on the VRAG. Quinsey *et al.* argue that exceptions in such cases are justified but "policy based" (because hospitals should have policies preventing the discharge of people who are making credible threats), and should therefore not be seen as reflecting on the validity of the instrument. But I am not sure of the importance of the terminology. From the point of view of distinguishing the different structured approaches, it seems safe to say that the advocates of actuarial methods are more inclined to stick to predictions made using data derived from groups of similar patients, while advocates of SPJ are more inclined to grant a role to clinical discretion.

This difference is related to a second distinction between actuarial and SPJ approaches. SPJ is designed to aid clinical management through attention to the items that the instrument comprises. Yes, the VRAG gains predictive accuracy in some samples by treating mental illness as protective, and Webster and Hucker's work suggests that, at least on the MacArthur sample, the HCR-20 could do the same. But the HCR-20 includes major mental illness and active symptoms among its items in part also because these variables can be addressed in treatment. Actuarial approaches are designed to aid clinical management too. But the emphasis in *Violent Offenders* is on their ability to point to the most appropriate placement, be this in conditions of high security, on an open ward, or in the community. When it comes to what treatment to offer in those settings, Quinsey *et al.* note how much is already known. "Criminogenic risk factors," they point out, including relationship instabilities and substance use, have been studied inside and outside the mental health field for years and don't require any particular form of risk assessment to be addressed. So do it.

There is a testable question here. Maden and Webster and Hucker both imply that treatment should be directed specifically at the dynamic items contributing to a patient's high score if that treatment is to reduce risk. Maden argues also that we can use changes in dynamic risk items, such as insight or

active symptoms of mental illness, to tell us whether risk has been reduced in an individual case. Designing a methodology that distinguishes the effectiveness of the more generic approach of Quinsey *et al.* from an item-based approach to the HCR-20 would be difficult, given that so many of the HCR-20 clinical items, including lack of insight, negative attitudes, and impulsivity, are likely to be addressed also by the models that Quinsey *et al.* recommend. But I don't think it is yet clear whether a group whose HCR-20 score has fallen is a group whose rate of violence has reduced or whether changes in the clinical variables account for any such reduction. And it would be nice to know.

Structured Instruments and Clinical Practice

These books also raise wider questions about the extent to which clinicians should be using structured approaches to assess risk as they go about their daily work. An initial inquiry might be, "Can structured assessments help us assess risk more accurately?" There is general agreement, here and elsewhere, that over the long term they have been shown to be more accurate than unaided clinical judgment. A subsequent question might be, "Does it matter which structured method I use?" Probably not, we are told. *Violence Risk* ends with a quiz. One true-or-false item stands out as potentially saving us all a lot of time: "When various contemporary risk assessment guides are pitted against one another in formal studies of predictive accuracy there tends to be no obvious 'standout' scale or device." The answer in the key at the back of the book is, True.

But the wider questions that address the details of clinical practice are more difficult to answer. "Will the structured risk assessment help me decide whether to discharge the patient?" is one, and it seems critical. There are three points to make here, I think. First, the accuracy, or inaccuracy, of the predictions that structured assessments allow seems important in assessing their usefulness: all things being equal, the more sure I am that something bad will happen if the patient is discharged, the more likely I am to keep that patient in the hospital. Maden is bullish on this point, stating that the accuracies involved here are on a par with those elsewhere in medicine. The consequences are so different that I am not sure we should apply the same standards that we apply for other medical tests, but I don't think Maden can be correct on this. A sensitivity of 73 percent at a specificity of 63 percent, one result reported from using the VRAG to predict violence (and towards the top end of all results achieved in psychiatric risk assessment), is substantially lower than the equivalent sensitivities of imaging tests and of most forms of cancer screening.[4]

The second point is that accuracy cannot be the whole story. If it were, actuarial prediction and SPJ would long since have replaced clinical judgment for predictions intended to cover periods lasting anything over a few days or weeks. If we are to put the absence of change down to something other than inertia or professionals' feeling threatened by new technology, we have to explain why clinicians have not adopted more generally what, after all, appears to be the best means going of predicting violence. My hunch is that it is partly because structured assessments, whether they generate a number or a classification into "high," "medium," and "low," give those clinicians information in a form that they find difficult to use. A doctor might worry that there is a risk of violence if the recently delusionally jealous husband returns to his partner, and worry also that her moving out will provoke him to violence. The same doctor may also know that the partner has been effective at limiting her husband's alcohol intake and persuading him to take his medication. So what to recommend? Clinicians make such judgments all the time. They are not just complex, but are inherently contingent: the right intervention in one circumstance turns out to be the wrong one in another, circumstances change, and not all circumstances can be controlled. Numbers and classifications can obviously provide a "background" against which these judgments are made. Beyond this, they may have a role, but it isn't immediately obvious how.

Psychiatrists do not assess risk in isolation from all of the other things they do when they manage a case. Risk to others is only one of many interacting considerations that go into a decision to allow a detained patient to spend time in the community, for instance, or to be discharged from the ward. No one has yet shown what should be the relationship between these decisions and the results of a structured assessment. There seems not to have been much at the time of the Ontario Survey by Quinsey *et al.*, the results of which are included in *Violent Offenders*. Patients with high VRAG scores were being allowed

to spend time in the community, usually, it appears, because they exhibited few symptoms. The findings are consistent with subsequent work by the same authors showing that actuarial scores are not significantly associated with tribunal decisions to transfer patients to lower levels of security or to the community.[5] Quinsey *et al.* and, somewhat more equivocally, the other authors under review see this as a problem. Structured assessments, they agree, should at least "inform" both placement within the hospital and any decision to discharge.

Whether other clinicians should also regard this as a problem depends on what is meant by "inform the decision." If by completing structured assessments clinicians will have available to them information that they would not otherwise have, or would have available to them the same information in a more reliable form (perhaps as a result of Webster and Hucker's "improved definition"), it seems hard to quibble. But a harder version of "inform the decision" would have a patient's ability to move to a predischarge ward or to the community depend on the results of a structured assessment falling below a given threshold. Here, I see little reason to doubt the impression left by Quinsey *et al.* that this would require a substantial reorientation of services. I assume also that this reorientation would not be limited to forensic services or discharge decisions. I don't know what the statistical relationship is at present between the results of an assessment using the HCR-20 in the emergency room and a subsequent decision to admit, but I doubt it is exact. It may not even be close.

One reason it may not be close is that when it comes to prevention, psychiatric services approach different types of violence differently. Structured assessments do not distinguish violence with different causes: an assessor using the HCR-20, for instance, is required to state only whether the patient is at low, medium, or high risk of violence, not that the risk derives from any particular source. But clinicians do make distinctions of this type. Anthony Maden writes that, "Not all acts of violence are equal, and we have to attach different value to the prevention of different acts." He notes that violence driven by psychotic symptoms is particularly important for mental health services. I think this is a complicated area: little attention is paid to the detailed etiology of a particular act when services are being blamed for allowing it to happen. But Maden is describing the views of many clinicians and, perhaps, many of those who design and fund what we do. If some forms of violence are more the province of psychiatry, should not the risk of these forms be a more important reason to admit?

Third, the knottiest problem of all. Can the results of using any structured risk assessment instrument, instruments that derive their validity from being tested on large groups of patients, ever be relevant to the management of an individual case? All of the books under review wrestle with this question—none of them very satisfactorily, to my mind. One of them concludes, within the space of five pages, that actuarial tools can inform the clinical assessment of an individual but that they should never be used in this way. But I am in no position to be critical. I don't know the answer. And I suspect I am not alone. The question of the relevance of data derived from groups to the management of an individual case has been debated, often with the greatest methodological sophistication,[6] since before any of these instruments were born.[7] In my view, there is little sign of a resolution.

One may take refuge in three statements. The first is that, for a clinician trying to manage a case, it should not be irrelevant that the patient scored 35 on the VRAG, or 33 on the PCL-R or a 0 on only 1 of 20 items on the HCR-20. The second is that this information is not dispositive. Risk of harm to others is only one of many factors that clinicians have to keep in mind as they develop their plans. And while there is a relationship between a score, or a category, on the one hand, and an individual's risk, on the other, that relationship is fuzzy. The third is that exactly how fuzzy depends in large part, and probably not in a way that statistical formulae can help most of us to describe, on the extent to which our patient and his circumstances resemble the patients and their circumstances in studies where the instruments have worked in the past.

Priorities, Thresholds, and the Future

I found few statements in these books that I would flat out argue with. But I am not sure that threat control override research has yet generated important and consistent findings, as one of them argues. And I wonder whether most of what matters in terms of history can be obtained from the patient, as Maden suggests. Records and other people are such obvious sources of potentially useful information that it would take some convincing for me not to use them.

In fact, I read the literature as suggesting that I continue to do so. One study showed that only 34 percent of collateral reports of violence were also reported by the subject.[8] More generally, studies that use multiple sources seem to report higher rates of violence.[9]

So what to do? Enough good research has been conducted, much of it by these authors, that the usual recourse to "more research needed" seems particularly inappropriate. If the clinician believes, with Quinsey *et al.*, that the risk of harm to others over the medium to long term is the primary criterion for discharge, they will use an actuarial instrument to measure that risk and apply a threshold to the score on that instrument to assist in their decision-making. They will know that, where the base rate of serious violence in the population is 3.6 percent in six months, as it was for patients in the CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness) study, a single act of violence can be prevented, even by an instrument as accurate as the VRAG, only by detaining 15 potential actors for the same period.[10] They may find this alarming, but if discharge is dependent solely on risk, it may be acceptable.

If, on the other hand, the clinician wishes to take advantage of whatever predictive accuracy is available but also believes that some of the important factors are particular to each patient and, hence, absent from any instrument, then that clinician will presumably be more sympathetic to an approach, such as SPJ, that will allow him to take some of those factors into account. In these cases, the complicated relationship between an individual's risk category and his own risk of offending, operating on an equally complicated relationship between his own risk of offending and any decision to discharge will mean that, in many cases, the results of the structured assessment of risk will not predict how the case is managed.

Finally, it seems unlikely that extended risk assessments will ever become the norm when mental health services come into contact with clients because the resource implications will simply be too great. Some kind of threshold, or filtering, will presumably then be used to divide "routine" cases from those who will receive a further evaluation. Structured approaches may have a role in such extended assessments, although perhaps more as ways of locating areas of concern than as final arbiters of what will happen. But their larger role may be in allocating cases to extended assessments in the first place. This, I would guess, is one point in the decision-making process that will receive extensive scrutiny when things go wrong, as these books show they will continue to go wrong, however well services are run. When that happens, it will help those services if they can show that their decision not to conduct an extended assessment was made rationally.

The technology described here does not yet, to my mind, take psychiatry to the point where clinicians can offer substantial improvements in public safety, over and above whatever we presently achieve, through the identification and clinical management of risky people. The limits to our ability to identify those who will offend mean that the lowest error rates presently achievable are still so high that admission to hospital solely to prevent violence, or preventive detention outside hospital to the same end, is accepted only when those being detained belong, as with sex offenders, to the most stigmatized groups in society. But that may change. The development of more accurate instruments, predicted here by Quinsey *et al.*, may permit error rates low enough for society to tolerate the detention of other groups. If that happens, the consequent upheaval in mental health services will presumably make the present disagreement between foxes and hedgehogs look like a very minor squabble.

References

1. Berlin I: The Hedgehog and the Fox. New York: Simon and Schuster, 1953
2. Leavitt J: Typhoid Mary. Boston: Beacon Press, 1996
3. Hopkins D: Princes and Peasants. Smallpox in History. University of Chicago Press: Chicago and London, 1983
4. Swets J: Measuring the accuracy of diagnostic systems. Science 240:1285–93, 1988
5. McKee S, Harris G, Rice M: Improving forensic tribunal decisions: the role of the clinician. Behav Sci Law 25:485–506, 2007
6. Hart S, Michie C, Cooke D: Precision of actuarial risk assessment instruments: evaluating the 'margins of error' of group v. individual predictions of violence. Br J Psychiatry 49(suppl):S60–5, 2007
7. Allport G: The psychologist's frame of reference. Psychol Bull 37:1–28, 1940
8. Mulvey E, Shaw E, Lidz C: Why use multiple sources in research on patient violence in the community? Crim Behav Ment Health 4:253–8, 1994
9. Lidz C, Banks S, Simon L, *et al*: Violence and mental illness: a new analytic approach. Law Hum Behav 31:23–31, 2007
10. Buchanan A: Risk of violence by psychiatric patients: beyond the "actuarial versus clinical" assessment debate. Psychiatr Serv 59: 184–90, 2008