Sixty-Six Years of Research on the Clinical Versus Actuarial Prediction of Violence

N. Zoe Hilton
Grant T. Harris
Marnie E. Rice
Mental Health Centre Penetanguishene

In their meta-analysis of clinical versus statistical prediction models, Ægisdóttir et al. (this issue) extended previous findings of statistical-method superiority across such variables as clinicians’ experience and familiarity with data. In this reaction, the authors are particularly interested in violence prediction, which yields the greatest support for actuarial models. In the past decade, actuarial prediction has continued to improve, but clinicians have not readily adopted these models, and new models have emerged that encourage reliance on unaided clinical judgment. Psychologists have made progress developing and disseminating actuarial risk assessments and should use the most accurate available measure suited to the task.

Grove, Zale, Lebow, Snitz, and Nelson (2000) reported that statistical prediction was about 10% more accurate than clinical prediction and was consistently superior across date and source of publication, type of judge (medical vs. psychological), general or task-relevant experience, type of data (e.g., interview results, psychological tests, trait ratings, behavioral observations, criminal record), and amount of data available. Statistical methods predicted forensic outcomes (e.g., criminal and violent behavior) especially well, mean effect size ($d$) of .89, but also fared well when predicting other outcomes.

Ægisdóttir et al. (2006 [this issue]) focus on outcomes of interest to counseling psychologists, including violence, and extend the findings to various statistical formulas, reliability and validity of the outcome criterion, and comparisons across clinicians’ familiarity with the setting, base rates, and the statistical formula itself. Our particular interest is violence prediction, which along with other criminal outcomes yielded the greatest superiority for statistical prediction (mean effect = .17). As Ægisdóttir et al. note, this effect size means that out of 1,000 predictions, statistical predictions accurately identify 90 more violent clients than do clinical predictions. Ægisdóttir et al. conclude that the overall advantage of actuarial methods is small but important (most conservatively, mean effect = .12) and encourage research on how to increase the application of actuarial methods. Thus, “when counseling
psychologists have familiarized themselves with available statistical formulas and prediction techniques, they should use those formulas and techniques” (Ægisdóttir et al., 2006, p. 370, emphasis added).

Furthermore, where counseling psychologists adjust existing statistical formulas, the adjustment itself should be actuarial—that is, based on empirical data regarding their own clients (especially, we would suggest, the base rate in their population). We agree with these conclusions, but drawing from our experience in training clinicians and counselors to use actuarial risk assessments for violence, we are concerned that some readers might take these conclusions as encouragement to adjust actuarial results on the basis of clinical intuition and in ways not supported by evidence. In addition, research and developments on violence risk assessment in the past decade since the most recent publication included in Ægisdóttir et al.’s meta-analysis suggest that the transition to reliance on actuarial methods is slower and more difficult than expected.

THE MEANING OF ACTUARIAL (VERSUS CLINICAL) PREDICTION

First, let us elaborate the distinction between actuarial and clinical, at least with respect to violence risk assessment in the psychological literature. Appraising the likelihood that an individual will be violent in the future entails two conceptually distinct tasks. The first is to select which personal or circumstantial characteristics to attend to. The actuarial method typically bases selection on one or more well-designed follow-up studies that identify which items are actually related to the outcome. This permits selecting an optimum set of items on the basis of incremental validity—that is, selecting the most powerful predictors first and then adding items only when they improve prediction. The past decade of research on violent recidivism indicates that nonclinical constructs (e.g., age, violent and criminal history) are important predictors of violent behavior. Clinical judgment bases selection on intuition, nonempirical experience, and one’s memory for empirical findings, and in assessing violence risk, clinicians attend to, or claim to attend to, psychiatric symptoms and diagnosis (Hilton & Simmons, 2001; Quinsey & Ambtman, 1979).

It is important to note here, however, that it is the method of selection rather than the items attended to that distinguishes clinical from statistical prediction. Indeed, empirical research has also established that many clinical constructs (e.g., psychopathy, conduct disorder, early childhood aggression) are central to optimal violence risk assessment. Some clinical constructs—most notably, symptoms or diagnoses of serious mental disorder—are empirically
unrelated or inversely related to subsequent violence in all clinical populations studied: civil psychiatric patients (Monahan et al., 2001), forensic patients and mentally disordered offenders (Bonta, Law, & Hanson, 1998; Harris, Rice, & Cormier, 2002), sex offenders (Hanson & Bussière, 1998), and violent offenders in general (Harris, Rice, & Quinsey, 1993). Thus, selecting psychotic diagnoses and symptoms as positive risk factors would embody clinical judgment, whereas a system, however counterintuitive, that treated such characteristics as protective factors would embody actuarial approaches. Because some of the best indicators require clinical skill to measure, accurately appraising violence risk is likely to remain a task for the clinician, but the place for the clinical judgment is within rather than outside actuarial tools.

The second task in risk assessment pertains to how risk factors are combined to render an interpretation. Leaving the combination rule unspecified represents unaided clinical judgment, as does relying on “gut-level” processes or permitting the use of idiosyncratic items. Conversely, combining risk factors using item weights derived from empirically established relationships with violent recidivism represents prototypical actuarial methods. Where empirical work shows that differential item weighting affords only small benefits, an actuarial system may weight all items equally (Harris et al., 1993; Hilton et al., 2004). Interpretation of final scores based on norms (the proportion of cases falling above or below a particular score) plus experience tables (the proportion of cases at a particular score that met the criteria for an operational definition of violent recidivism within a specified opportunity) represents prototypical actuarial methods.

TRENDS EVIDENT IN VIOLENCE RISK ASSESSMENT IN THE PAST 10 YEARS

1. Actuarial Assessments Seem to Be Increasingly More Accurate

More research demonstrating that the outcome of unstructured clinical assessments left a great deal to be desired seemed to be overkill. That horse was already dead. (Monahan et al., 2001, p. 7)

Grove et al. (2000) considered the literature up to 1988 and concluded that actuarial approaches were about 10% more accurate than clinical approaches, with a trend toward a greater difference for forensic predictions. Ægisdóttir et al. (2006), considering the literature up to 1996, concluded that actuarial approaches were 13% more accurate and, in the case of predictions about future violent or criminal behavior, 17% more accurate. More recently, Hanson and Morton-Bourgon (2004) reported that for sexual violence,
actuarial assessments had an effect size 88% larger than did clinical judgment. We believe this trend is because of empirical approaches’ discovering better predictors, better ways to weight and combine variables, and better ways to improve reliability. Clinical judgment, being much less tied to empiricism, improves little over time. This coincides with the finding that clinical experience and “expertise” add little to the accuracy of clinical judgment.

2. Actuarial Violence Risk Assessments Are Available but Not Always Used

The decade after the most recent article included in Ægisdóttir et al.’s (2006) meta-analysis saw the advent of prototypical actuarial systems for all male offenders, including mentally disordered and convicted offenders (Harris et al., 1993; Quinsey, Harris, Rice, & Cormier, 2006); for sex offenders (Hanson, 1997; Quinsey et al., 2006); and for wife assaulters (Hilton et al., 2004). Despite the advent of these actuarial methods, clinicians and counselors did not abandon unstructured clinical judgment and embrace statistical prediction. Instead, there was a desire to adopt a middle ground. For example, clinicians use actuarial tools as only part of their assessments, combining them in various ways, yet the evidence suggests that combining assessments does not increase the accuracy of prediction and can reduce it (Seto, 2005).

In examining a tribunal’s decisions about the detention or release of forensic patients, Hilton and Simmons (2001) illustrated the deleterious effects of trying to assimilate actuarial assessments into normal clinical practice. They conducted the study at an institution where a prototypical actuarial assessment for violence risk (the Violence Risk Appraisal Guide [VRAG]; Harris et al., 1993) had been available for several years. The tribunal was legally required to make release decisions primarily on the basis of violence risk. In some cases, the VRAG was available to the tribunal in a hospital staff report, and in other cases, the VRAG was not scored until after the tribunal hearing. In all cases, oral testimony about risk was provided by a senior clinician, usually a psychiatrist. The information used to score the VRAG was entirely drawn from the patient’s institutional file, also available to the psychiatrist. Hilton and Simmons found that neither psychiatrists’ recommendations nor tribunal decisions were associated with the actuarial VRAG score; in the risk of violent recidivism, there was no difference between patients detained and those transferred to lower security or released.

Recidivism was significantly related to actuarial score but not to psychiatrists’ opinions. Access to the actuarial score did not improve their performance and tended to worsen it, underscoring Ægisdóttir et al.’s (2006) report that access to statistical information did not improve clinicians’
predictions (and tended to worsen them). This finding is consistent with our experience that clinicians deliberately base their advice about violence risk on information other than actuarial scores, believing their task is to provide a different perspective, which in Hilton and Simmons’s (2001) study was more strongly related to psychiatric symptoms than to the empirical predictors of subsequent violent behavior.

If forensic decisions mirror unaided clinical judgment and not actuarial assessments after a well-validated actuarial system has been available for a decade, we must concede that the mere availability of actuarial methods—and the careful analysis and dissemination of their consistent advantage—is fundamentally insufficient for their adoption. We do not believe that this apparent resistance is mainly conscious and deliberate obstinacy or an inability to comprehend the statistical evidence. Experimental studies conducted at our institution revealed that forensic clinicians can make proficient hypothetical decisions using actuarial scores (Hilton, Harris, Rawson, & Beach, 2005). Rather, improving forensic practice must depend on something else. But what? One attempt to find some middle ground has been structured clinical risk assessment.

3. One Search for Middle Ground Meant Reliance on Clinical Judgment

Efforts to moderate the statistical and clinical positions saw the emergence of structured professional judgment (SPJ; reviewed in Douglas, Yeomans, & Boer, 2005), and several guides have been created for violence in general (Historical Clinical Risk–20; Webster, Douglas, Eaves, & Hart, 1997), sexual violence (Sexual Violence Risk–20; Boer, Hart, Kropp, & Webster, 1997), and spousal violence (Spousal Assault Risk Assessment Guide; Kropp, Hart, Webster, & Eaves, 1995). These guides represent empirical approaches in that some items were selected on the basis of the follow-up literature and that some are empirically robust violence risk factors (e.g., criminal history, substance abuse, psychopathy, violation of conditional release). On the other hand, many were not identified as risk factors in follow-up research and seem to have been selected on the basis of clinical experience and intuition (e.g., child abuse victimization, suicidal or homicidal ideation, lack of insight). Some items (e.g., severity of violence, psychotic symptoms or diagnosis) were offered as positive risk factors despite contradictory empirical evidence, perhaps because clinicians erroneously believed them to be violence risk indicators in clinical populations.

We suggest that a list of factors gleaned partly from empirical follow-up literature and partly from nonempirical clinical opinion falls between prototypical clinical and actuarial methods. Although these SPJ schemes did not use
statistical optimization methods in their development, subsequent evaluations 
have reported that total scores predict the outcomes of interest (de Vogel, de 
Ruiter, van Beek, & Mead, 2004; Douglas et al., 2005; Hanson & Morton-
Bourgon, 2004; Hilton et al., 2004; though see Kropp & Hart, 2000), although 
not usually as well as actuarial instruments and sometimes no better than 
unaided clinical judgment. In clinical practice, however, SPJ schemes dis-
courage using total scores and encourage users to incorporate additional 
idosyncratic risk factors where necessary on the basis of unaided clinical 
judgment. Users are directed to assign idiosyncratic weights (focusing on 
“critical items”) and to render a final trichotomous interpretation (low, 
moderate, or high risk) in the absence of specific criteria. Ægisdóttir et al. 
(2006) correctly point out that user-friendly models often mirror the cog-
nitive processes of clinical judgment: Freedom to select items and to modify 
item weights, and rendering an interpretation in the absence of any criteria 
for doing so, clearly represents prototypical clinical judgment.

SPJ schemes were greeted with optimism because they “provide[d] a 
sophisticated and flexible framework within which to exercise professional 
discretion” (Kropp & Hart, 2000, p. 116). Yet interrater reliability of the 
final assessment has been poor (de Vogel et al., 2004; Douglas et al., 2005; 
Kropp & Hart, 2000), much lower than reported for prototypical actuarial 
methods (Harris et al., 2003; Hilton et al., 2004) and even for the simple 
item total scores on the SPJ schemes. Consistent with Ægisdóttir et al.’s 
(2006) conclusions that logical rules are no better or worse than clinical 
methods, we conclude that this arises from the heavy reliance of SPJ on 
clinical judgment. Ægisdóttir et al. observed no improvement in clinical 
judgment accuracy in the first 56 years, and it appears that current attempts 
to find a clinician-friendly middle ground for violence risk assessment have 
not resolved this problem.

4. Expertise Is Gained From Follow-Up Experience: 
Psychologists Can Learn From Clients

As mentioned earlier, follow-up research gives actuarial risk assessment 
the advantage of empirical feedback, which clinicians rarely get. That is, 
clinicians routinely make predictions of violence but rarely are exposed to 
any information about the success or failure of the predictions, especially 
when the predictions determine actual opportunity. Counselors working 
with victims of violence might be an exception, especially those working in 
crisis centers for women victimized by their partners. Such clients returning 
after repeated assaults might provide some information about prediction 
accuracy. Moreover, the victims themselves, exposed to many samples of 
their partner’s behavior, might be in a position to learn about their own
predictions. Emerging empirical research supports the notion that victims’ perceptions of the likelihood of future violence are a good indicator of actual risk (Cattaneo, in press; review in Hilton & Harris, 2005) and were empirically selected as an item in an actuarial assessment of domestic violence risk (Hilton et al., 2004). Cattaneo (in press) reported that both victims’ and advocates’ predictions were significantly associated with variables supported by the existing research as predictors of interpartner violence (e.g., substance abuse, general violence, children, prior physical violence).

WHITHER GOES ACTUARIAL AND STATISTICAL PREDICTION OF VIOLENCE?

In reviewing developments in the past 10 years, our point is not so much to rehash the relative value of existing schemes; SPJ sometimes seems to work, just as both actuarial and clinical judgment seem to work overall. Rather, we observed a change in the field’s position after the prototypical actuarial methods for violence risk assessment became widely available. Instead of simply adopting them (as Ægisdóttir et al. [2006] imply would be expected), many forensic clinicians gravitated to something else, ostensibly a moderate position but in fact clinical judgment in new clothing. Ægisdóttir et al. (2006) have performed an extremely valuable service in examining and communicating the size, nature, and limits of the superiority of actuarial methods, but recent experience in the field of violence risk assessment suggests that their work will be insufficient to produce measurable improvement in practice. Our position has been described as extreme (including by Ægisdóttir et al.), yet the apparent middle ground has, in our opinion of the empirical evidence, proved to be a stronghold for unaided clinical judgment. Why is practice so resistant to Ægisdóttir et al.’s advice to use actuarial methods?

Many arguments against the value of actuarial methods in general (Grove & Meehl, 1996) and actuarial violence risk assessment specifically (Quinsey et al., 2006) have been examined in detail. Ægisdóttir et al. (2006) have cogently summarized the controversy, and a complete treatment of all objections is beyond our scope here. We wonder whether psychologists and other clinicians fear the loss of prestige in their clinical role. When evaluating a behavior therapy program, Hilton and Simmons (1999) observed that clinicians (against the rules of the program) sometimes assigned points in a behavior therapy program without observing patients’ behavior, and we suggest that clinicians can develop a sense of confidence and even pride in their familiarity with their clients’ behavior and potential such that they occasionally overlook the need to seek data to disconfirm their perceptions.
Ægisdóttir et al. show that clinicians actually made more accurate predictions when they were not familiar with the source of the data on which they based their predictions. Indeed, some psychologists select clinical work as a field of study on the mistaken idea that it is primarily nonquantitative, requiring outstanding interpersonal skills alone.

On a positive note, the past decade has seen progress in the transfer of knowledge from actuarial research to the training of counselors and other professions. Psychologists have largely contributed to this trend. From our experience, a book we wrote about actuarial risk assessment (Quinsey et al., 1998) has generated hundreds of requests for training and several hundred requests from clinicians about precise scoring of items. The Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004) was intentionally developed for counselors as well as frontline police officers assessing the risk of men’s violence against their female partners, and our work in progress shows that ODARA training significantly improves psychologists’ and other counselors’ ability to assess violence risk accurately and increases the association between actuarial risk and police officers’ decisions to release or detain men charged with spousal assault.

CONCLUSIONS

First, they say you’re wrong and they can prove it. Second, they say you’re right, but it’s not very important. Third, they say you’re right and it’s important, but they’ve known it all along. (Kettering, cited in Quinset et al., 2006, p. 12)

Efforts to incorporate actuarial methods into clinical practice regarding violence risk assessment are at Stage 2. Ægisdóttir et al.’s (2006) contribution helps us proceed from there. Psychologists are often the only professionals in counseling settings to be trained in statistical skills, and it behooves our profession to use the most accurate available measure for the task. We need skilled clinicians to develop, implement, and evaluate potent interventions for reducing the risk of violence. In the meantime, actuarial methods provide psychologists with the most accurate means to make long-term evaluations of violence risk and to apportion intervention resources aimed at lowering this risk.

REFERENCES


