Psychiatry, Psychology and Law, 2017
Vol. 24, No. 2, 292–301, http://dx.doi.org/10.1080/13218719.2016.1197817

Covert and Implicit Influences on the Interpretation of Violence Risk Instruments

Stephane M. Shepherd[a] and Danny Sullivan[a,b,c]

[a]*Centre for Forensic Behavioural Science, Swinburne University of Technology, Melbourne, Australia;* [b]*Department of Psychiatry, University of Melbourne, Australia;* [c]*Victorian Institute of Forensic Mental Health (Forensicare), Melbourne, Australia*

Forensic mental health practitioners are frequently asked to estimate the risk of future violence. Legal decisions concerning the sentencing, management and disposition of offenders often rely on the advice of such testimony. The burgeoning use of violence risk instruments in these settings undoubtedly injects a level of scientific rigour into forensic evaluations for courts and tribunals. Yet scrutiny of the inherent limitations of both risk instruments and the inferences and formulations drawn from them are often veiled by the discipline's endorsement for such approaches. Misconceptions about the validity and dependability of present-day risk assessments and expert infallibility persist. The furtive influences that shape both the (mis)interpretation and miscommunication of risk instruments in legal settings necessitate discussion.

Key words: expert testimony; forensic assessment; forensic psychiatry; violence risk assessment; violence risk instruments.

Introduction

The historical progression from impressionistic clinical predictions of dangerousness to the utilisation of violence risk instruments is well documented (Webster, Douglas, Eaves, & Hart, 1997). Noteworthy court decisions (*Baxstrom v. Herald,* 1966) and ensuing scientific studies (Steadman & Cocozza, 1974) drove this evolution, which in essence juxtaposed opposite ends of the risk prediction spectrum: elementary clinical judgement and mechanical actuarial risk instruments. Motivation for the development of actuarial instruments intensified after empirical research demonstrated the inclination for clinicians to over-predict violence risk (Monahan, 1981). This movement represented a paradigm shift in the forensic discipline. 'Error-prone' clinical risk adjudication became anathema to objective assessment, while statistical forecasting – ostensibly free from heuristics and illusory correlations – became the process de rigueur. Clear-cut risk items on actuarial tools provided a standardised scientifically-informed platform for risk evaluation, enhancing both rater consistency and transparency. Actuarial measures however were not without criticism; their inflexible constitution and usurping of clinical decision-making engendered apprehension in the literature (Doyle & Dolan, 2007; Hart, 1998).

Despite calls to replace clinical judgement entirely with actuarial tools (Quinsey, Harris, Rice, & Cormier, 2006), middling

Correspondence: Stephane M. Shepherd, Centre for Forensic Behavioural Science, Swinburne University of Technology, 505 Hoddle Street, Clifton Hill VIC 3068 Australia. Tel.: (03) 9947 2606. Email: sshepherd@swin.edu.au

© 2016 The Australian and New Zealand Association of Psychiatry, Psychology and Law

alternatives like 'structured professional judgement' (SPJ) enabled clinical discretion to be anchored by a checklist of static and dynamic evidence-based risk factors drawn from the scientific literature (Heilbrun, Yasuhara, & Shah, 2010). The positivistic nature of a predetermined set of risk factors – a remnant of the actuarial approach – is thus complemented with a clinician's contextual oversight. Though often requiring more time to administer compared to actuarial instruments, SPJ approaches are widely utilised by forensic practitioners (Hurducus, Singh, de Ruiter, & Petrila, 2014; Khiroya, Weaver, & Maden, 2009). Later models of risk assessment now incorporate supplementary risk management frameworks to inform risk reduction strategies – such instruments are often included within broader risk-assessment paradigms such as the Risk-Needs-Responsivity (RNR, see Andrews & Bonta, 2010) framework where the evaluation of risk is one of three phases in offender management.

Despite these advancements in violence risk prediction, the nomothetic nature of risk instruments is often at odds with notions of individualised, case-specific judgements in legal practice. The risk-prediction testimony of mental health practitioners providing forensic evaluations for courts and tribunals is generally expressed in probabilistic terms, although it will 'almost always be used to justify all or nothing decisions' (Mullen & Ogloff, 2009, p. 1994). Navigating this contradictory terrain presents several professional and ethical challenges for forensic practitioners. For this reason it is important that both forensic evaluators and the judiciary are cognisant of the existing limitations of violence risk instruments, which at times enjoy unchallenged medico-legal approval. In response, this paper aims to traverse the often overlooked shortcomings of violence risk instruments, from *validation* through to *interpretation*. The limitations pertaining to instrument utility, validation procedures, and applicability to unique offender circumstances and populations are summarised, and common practitioner biases and heuristics affecting risk evaluations are additionally canvassed. Emphasising clinical prudence, this commentary seeks to challenge the oftentimes uncritical application of violence risk instruments in legal settings. The notions surveyed in this article are derived from the combination of clinical experience and a thorough synthesis of the international forensic mental health academic literature pertaining to violence risk assessment and expert testimony.

Lies, Damned Lies and Statistics

Traditionally, risk instruments are validated on their capacity to predict recidivism alone, regardless of their explicit purpose (i.e. to predict, estimate, manage risk, etc.). Instruments like the Violence Risk Appraisal Guide (Quinsey, Harris, Rice, & Cormier, 2006) are comprised of largely static items with the aim of *predicting* recidivism, while other instruments, such as Historical Clinical Risk Management-20 (HCR-20; Douglas, Hart, Webster, & Belfrage, 2013) and the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) are designed to *estimate* a client's level of risk. Although arriving at risk estimations through differing processes, HCR-20 and the LS/CMI ostensibly aim to *manage* a client's level of risk with an end goal of *reducing* risk by targeting identified dynamic risk items for intervention (Doyle & Dolan, 2002). Evidence for the predictive utility of risk instruments is relatively robust, although evidence for the capability of risk instruments to effectively manage and ultimately reduce risk remains elusive. It is necessary to explore both outcome measures in greater detail.

Myriad investigations have demonstrated associations between risk instruments and future offending. However, forensic validation commentary often overestimates a risk instrument's capacity to predict recidivism, presumably galvanised by comparisons to 'flip-of-a-coin' clinical prognoses. This

enthusiasm materialises from the body of literature underlining the predictive advantage of risk instruments over unstructured clinical judgement (UCJ); the catchphrase 'superior to unstructured clinical judgement' is often equated with 'predictive accuracy'. It is problematic for the field if lower range predictive indices are not appropriately challenged simply because of their marginal improvement beyond clinical discretion.

Validation studies will often determine the efficacy of a risk instrument based on its ability to differentiate between reoffenders and non-reoffenders. This procedure is commonly explored using receiver operator characteristic (ROC) analysis. In ROC analysis, an *area under the curve* (AUC) index is calculated. So widespread is the use of ROC analysis in violence risk validation research that it has become the status quo – risk instruments essentially 'live or die' by the AUC index. In the literature, AUC scores of 0.75 are considered robust (Dolan & Doyle, 2000), indicating that there is a 75% chance that a randomly-selected recidivist will receive a higher score on an instrument compared to a randomly-selected non-recidivist. The majority of validation studies for widely-used risk instruments has demonstrated AUC scores below this level. In real terms, this may be an underestimation, given the propensity for authorship bias (Singh, Grann, & Fazel, 2013). Moreover, ROC analysis has a number of shortcomings. While the AUC index is a favoured indicator due to its resistance to base rate fluctuations (Babchishin & Helmus, 2015; Singh, 2013), it is not a true reflection of predictive accuracy. Furthermore, ROC analysis cannot meaningfully ascertain the forecasting ability of instruments comprising modifiable items (Falzer, 2013). Determining an instrument's positive and negative predictive values can provide a stronger indication of an instrument's predictive capacity (Singh, 2013; Szmukler, Everitt, & Leese, 2012). However, supplementary analyses identifying how changes in dynamic factors redefine both level of risk and future behaviours would bolster validation evidence. Such indicators are underused in the forensic field and warrant inclusion in validation studies so that this information can be routinely provided in court.

The risk prediction validation literature is riddled with methodological errors, such as authorship partiality (Singh et al., 2013), interpretational inconsistencies across studies (Singh, Desmarais, & van Dorn, 2013), incompatible cross-validation methods (Rossegger et al., 2013) and a common misunderstanding of the use of the AUC index (Singh, 2013). Discrepancies in replication methodology are unavoidable in the forensic field, where sustained access to various at-risk populations is often unfeasible; thus there are often numerous confounders when compared with the methodological exactitude of instrument construction samples. For the sake of expediency, deviation from the user recommendations underlined in risk instrument manuals is commonplace in validation studies. In many respects, we must 'take the good with the bad' methodologically to progress the field, given the unpredictable and sometimes inaccessible nature of forensic samples. However when nine out of ten risk instrument studies are misinterpreting AUC values (Singh et al., 2013), it is incumbent on the discipline to appropriately revisit its evaluative processes. If validation endeavours focusing on risk prediction continue, greater levels of methodological sophistication are warranted. There is a dearth of meaningful research exploring the influence of dynamic risk factors (risk state) on a client's level of risk over time and how this may vary with consideration to their risk propensity (risk status). In light of these outstanding issues, the predictive triumphalism found in many validation studies is potentially overstated.

How then should we approach the appraisal of risk instruments? It has been suggested that tools are of little worth if unable to *reduce* a client's level of risk. This elicits concern from several commentators who point to the unavailability of scientific

evidence for genuine risk reduction attributable to risk instruments (Nielsson, 2013; Wand & Large, 2013). In contrast, Allnut et al. (2013) posit that 'violence risk assessment is not an intervention and conducting an assessment of violence risk does not change the outcome' (p. 734). From this perspective, risk instruments – for example, the *Level of Service* instruments[1] – inform therapeutic approaches (via the RNR model perhaps) but are not actual interventions in isolation. In any case, validation difficulties arise for instruments which aim to manage risk rather than solely predict recidivism. Validating an instrument *with a view to risk management* by ascertaining its predictive capacity for recidivism is disingenuous – an instrument designed to *inform* risk management decisions should be evaluated on its ability to do just that. At the same time, validating the same instrument by the success of follow-on rehabilitative efforts is potentially too 'downstream'. This option is particularly complicated by an observed disconnection between identifying risk/needs and developing interventions based on those needs (DeMatteo, Hunt, Batastini, & LaDuke, 2010; Singh, Desmarais, et al., 2014). It could be argued that contemporary violence risk instruments currently exist in an evaluative vacuum. The field may need to consider novel and more relatable frameworks for the validation of contemporary instruments. With this in mind, clinicians must be particularly cautious when communicating risk-instrument information. As *clinical* aids, risk instruments may help to prioritise treatment goals and tailor management strategies – but research on whether they enable improved clinical decision-making beyond prediction is inadequate. The task for clinicians is to not fall into the trap of overvaluing the predictive properties of a risk instrument when formulating a risk opinion.

Special Cases

The nomothetic composition of violence risk instruments refers to their generalised, population-based criteria. Contemporary risk instruments are typically normed on the correlates of recidivism for cohorts of offenders from specific localities (actuarial) or they are assembled through a process of clinical endorsement and item selection from the violence risk literature (SPJ). Either way, the application of aggregate data to an individual is at odds with legal notions of individualised justice. A common objection is that group statistics may not apply to an individual when formulating risk, channelling the ecological fallacy. Realistically, individuals can deviate from instrument construction samples in several ways, threatening the generalisability of an instrument. This point is especially pertinent for actuarial instruments, which encompass a finite set of factors with *predetermined weightings*. This method fails to embrace important idiosyncrasies that are not measured by the instrument, including *inter alia* age and cultural, attitudinal, contextual and protective factors. These counterfactual elements may have a marked influence on risk for specific individuals and groups, an influence which is disproportionate and may have far greater bearing on risk that the general factors incorporated into nomothetic instruments.

The applicability of instruments to populations not widely represented in construction samples (such as ethnic minorities, females and unique offenders) is equivocal. Several key studies have found that predictive validity estimates for minority and/or female offenders are weaker compared with white male offenders (Edens, Campbell, & Weir, 2007; Holtfreter & Cupp, 2007; Shepherd, Luebbers, & Dolan, 2013; Shepherd, Luebbers, Ferguson, Ogloff, & Dolan, 2014; Singh, Grann, & Fazel, 2011). Recent studies from the United Kingdom also suggest that the presence of psychopathy drastically reduces the predictive validity of risk assessment instruments (Coid, Ullrich, & Kallis, 2013), notwithstanding that this is a group which has historically been singled out for preventive detention due to a strong

perception of escalated risk (Mullen, 2007). Meta-analytic studies assessing a range of risk assessment instruments demonstrate greater reliability for the prediction of low-risk categories, but a low level of reliability for assessments of high risk, leading Fazel and colleagues to contend that

> these tools are not sufficient on their own for the purposes of risk assessment. In some criminal justice systems, expert testimony commonly use scores from these instruments in a simplistic way to estimate an individual's risk of serious repeat offending. However, our review suggests that risk assessment tools in their current form can only be used to roughly classify individuals at the group level, and not to safely determine criminal prognosis in an individual case. (Fazel, Singh, Doll, & Grann, 2012, p. 5)

Such evidence has not, however, clearly led to reluctance to deploy risk assessment tools in support of preventive detention for sexual and violent offenders in numerous jurisdictions. Another key issue is the paucity of violence risk factor evidence for outlier groups, complicated by the feasibility of accessing such information. For example, offenders motivated by political or religious extremism may pose a significant risk of violence but not be assessed as such by risk tools. There the salient variables may relate to fervour and other variables not easily determined even by seasoned assessors. The narratives of radicalised terrorists often indicate a process taking place over some months, during which the level of risk alters markedly. However the markers of altered risk do not correspond to the same risk factors identified in the empirical literature for violence, and the form of violence may also differ (Dalgaard-Nielsen, 2010). Furthermore, the variegated motivations of those at risk for violent extremism, combined with the rarity of such events, renders risk prediction exceptionally complex and prone to oversimplification. Efforts have been made to develop such tools, e.g. the Violent Extremism Risk Assessment – Second Edition (VERA-2; Pressman & Flockton, 2010), but they have not been extensively validated.

First-time offenders pose a specific difficulty, particularly if they are being assessed prior to treatment. If treatment is considered to work then post-treatment risk assessment may have quite different outcomes depending on progress and the passage of time (which for younger offenders is accompanied by possible maturation, a factor that is often significant). Some offenders benefit markedly from treatment, and others do not; however, it is likely that treatment outcome is of marked relevance to prognosis and recidivism risk, although risk assessment prior to treatment interventions does not reflect this (Falzer, 2013).

These illustrations place the present-day estimation of risk, for those who do not conform to instrument norms, in a void: we are bereft of scientific instruction. On the one hand, tools that do not accommodate such populations may be unsuitable, while on the other, the alternative – intuition – may be similarly inappropriate. In each scenario, clinicians are forced to make an inferential leap when formulating a probabilistic estimate of risk in the absence of robust statistical evidence to inform such an estimate. This is of particular concern where SPJ instruments are utilised. An SPJ projection can still sustain many of the subjective biases that weaken unstructured clinical judgement (UCJ). Despite the availability of salient variables for consideration there is no defined logic about how the instrument's categorical overall risk rating (High, Medium, Low) is formulated. Research has identified significant discrepancies in both clinical non-numerical risk estimates (Hilton, Carter, Harris, & Sharpe, 2008) and violence outcome in clients rated as high risk (Singh, Fazel, Gueorguieva, & Buchanan, 2014). Furthermore, it is alleged that clinicians often forgo their discretionary option, choosing to score SPJ instruments in an additive, mechanical fashion (Hanson & Morton-Bourgon, 2009).

Ostensibly, the flexibility that exemplifies SPJ also renders the approach vulnerable to misuse and, in some cases, greater imprecision (Hanson & Morton-Bourgon, 2009). Indeed, these challenges should not preclude the use of SPJ instruments in court proceedings; however, it is incumbent on health professionals giving evidence to clearly articulate the inherent limitations of using such an approach, particularly with unique offenders with scientifically unidentified propensities for future violence. We would contend that it is also incumbent on the trier of fact not just to acknowledge the limitations of such instruments and proceed to use the evidence, but to reflect on whether such instruments lead to conclusions which accord with the legal standard of proof being applied.

Cognitive Biases

Risk instruments add a level of structure and transparency to estimations of violence. A range of clinician biases may be minimised as a result. However, the process is not immune to harmful subjectivity and prejudices – not just in the course of assessment but also in the interpretation of results (Dror, 2011; Neal & Grisso, 2014). Inter-rater reliability may differ depending on who commissions the report. Research has indicated that risk-instrument scores are susceptible to adversarial allegiance (Chevalier, Boccaccini, Murrie, & Varela, 2015; DeMatteo et al., 2014; Murrie, Boccaccini, Johnson, & Janke, 2008; Murrie et al., 2009). This phenomenon has also been shown in civil medico-legal assessments (Large & Nielssen, 2001), and to a lesser extent in criminal medico-legal assessments (Nielssen, Elliot, & Large, 2010). Some expert witnesses may gain a reputation for providing opinions which exhibit consistent tendencies. In an adversarial legal system, such perceived bias may indeed be the reason for the expert being retained for an opinion. Not only may witnesses exhibit subjective biases, but their repeated retention by the same lawyers or firms may also provide a subtle or unconscious bias to satisfy those who commission the reports. The ethical duties of a lawyer, within constraints, are to act in the best interests of their client. It is not a breach of their ethics to seek an advantageous report, and a good lawyer will no doubt preferentially seek a report which they suspect may benefit their client.

The mode of communication of results is critical. The assessor requires not only a sophisticated understanding of the assessment guidelines and statistical properties of any instruments used but also the capacity to explain them clearly to the trier of facts. This is a requirement of most codes of conduct for expert witnesses and has a clear pedigree in common law (*Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993; National Justice Compania Naviera SA v. Prudential Assurance Co Ltd (The Ikarian Reefer), 1993*). A range of cognitive biases may influence the presentation and understanding of risk assessments in medico-legal settings. As described eloquently and succinctly by Tversky and Kahneman (1974), the presentation of data influences the choices made, although decision-makers are often unaware of their biases.

Information framed in particular forms influences decisions strongly. For the medico-legal presentation of risk-assessment information this may be reflected in idiosyncratic understandings of the meaning of terms such as *high* and *low*, which refer not only to a comparison to another notional group such as a norm but also to the base rate of the behaviour in question. Is the same sort of violence being referenced, or another (lesser or greater) form? Is the risk to be considered relative to categories, for example, of *men*, *men of the same age*, *men of the same sociocultural background*? Is the context a jurisdiction with gross rates of violence or one in which the base rate is very low? It can be seen that terms such as *high*, *moderate* and *low* are inseparable from both the comparator group and the context.

Furthermore, such prognostications return us to the days of 'dangerousness', when risk

was considered akin to an individual attribute or trait (Mullen, 2007). The outcome of risk assessment becomes increasingly inaccurate with the passage of time from the assessment, and hence, in clinical practice, frequent reassessment is required. Yet such communications in a medico-legal context may affect events years in the future, such as the duration of a sentence or other restrictive conditions (McSherry, 2004). Prognostications are seen as indelible, notwithstanding a range of future contexts which are predominantly unforeseeable by the assessor. Future contexts involving not-yet-known associates, relationships, states of intoxication or prevailing mood states may be critical in future situational aggression or sexual violence. The more remote from the time of assessment, the less capable is the assessor to determine that their prognostication will be valid in such contexts.

Tversky and Kahneman (1974) describe the availability bias as reflecting the implicit influence of recent cognitions upon decision-making. For those undertaking risk assessment, recent events impact upon the assessment. A patient who suicided; a traumatised victim recounting the effects of a crime; a run of offences which are similar: these events will impact upon a risk assessor. Such events may make them more blasé or more cautious, or may skew their scenario planning. Notwithstanding the rigour of assessment guidelines, the subjectivity of the assessor is influenced by their own context. It is valid to ask about the assessors' experience with victims, or how they might account for their potential biases.

Finally, decision-making is influenced by anchoring effects. This term refers to the fact that the first facts offered will tend to influence final results. Interestingly, when risk assessment is sought by the court, the preliminary arguments to establish the eligibility of the offender for risk assessment focus upon the egregiousness of their history. Any discussion of risk assessment is preceded by the history of offending. Thus the argument is anchored in the heritage of the offender as a person with offending tendencies. Risk assessment is never undertaken as an abstract task, devoid of context. The anchoring effect of commencing proceedings with narratives of offending may set the tone for a bias towards estimation of higher risk.

For the trier of fact, maintaining imperturbable neutrality must be difficult. For the assessor, to extricate themselves from these biases requires a remarkable degree of self-awareness, humility, openness and flexibility. These are not necessarily attributes sought in the expert witness. More often experts are selected for their capacity to 'stick to their guns' and defend their assessments. The characteristics of an effective expert witness may conflict with his or her capacity to offer the frank, flexible and individualised assessment which is required, or to be circumspect.

Courts in the positivistic tradition seek robust responses rather than ambiguity. They do not prize equivocation. There are pressures, markedly in the adversarial system, which seek polarised and irrefutable, staunch opinions. It is contended that in this domain, the nuance of risk assessment, its errors and shortcomings, may be glossed over in order to satisfy the needs of the parties or legal system. While giving evidence about risk need not be unethical mental health practitioners should be both knowledgeable and cautious when they provide risk assessments to the courts.

Note

1. A systematic assessment of offender risks and needs (see Andrews et al., 2004).

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

Allnutt, S. H., Ogloff, J. R., Adams, J., O'Driscoll, C., Daffern, M., Carroll, A., Nanayakkara, V.,

& Chaplow, D. (2013). Managing aggression and violence: The clinician's role in contemporary mental health care. *Australian and New Zealand Journal of Psychiatry, 47*(8), 728−736. doi:10.1177/0004867413484368

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Newark, NJ: Matthew Bender.

Andrews, D.A., Bonta, J., & Wormith, S.J. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto, Ontario, Canada: Multi-Health Systems.

Babchishin, K. M., & Helmus, L. (2015). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world data. *Behavioral Research Methods*. Advance online publication. doi:10.3758/s13428-015-0627-7

Baxstrom v. Herald, 1966. 383 U.S. 107

Chevalier, C. S., Boccaccini, M. T., Murrie, D. C., & Varela, J. G., (2015). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior, 39*(3), 209−218. doi:10.1037/lhb0000114

Coid, J. W., Ullrich, S., & Kallis, C. (2013). Predicting future violence among individuals with psychopathy. *The British Journal of Psychiatry, 203* (5), 387−388. doi:10.1192/bjp.bp.112.118471

Dalgaard-Nielsen, A. (2010). Violent radicalization in Europe: What we know and what we do not know. *Studies in Conflict & Terrorism, 33*(9), 797−814. doi:10.1080/1057610X.2010.501423

Daubert v. Merrell Dow Pharmaceuticals, Inc. 1993., 509 U.S. 579

DeMatteo, D., Edens, J. F., Galloway, M., Cox, J., Smith, S. T., & Formon, D. (2014). The role and reliability of the Psychopathy Checklist-Revised in U.S. sexually violent predator evaluations: A case law survey. *Law and Human Behavior, 38* (3), 248−255. doi:10.1037/lhb0000059

DeMatteo, D., Hunt, E., Batastini, A., & LaDuke, C. (2010). The disconnect between assessment and intervention in the risk management of criminal offenders. *Open Access Journal of Forensic Psychology, 2*, 59−74.

Dolan, M., & Doyle, M. (2000). Violence risk prediction. *The British Journal of Psychiatry, 177* (4), 303−311. doi:10.1192/bjp.177.4.303

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk for violence − user guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Doyle, M., & Dolan, M. (2007). Standardized risk assessment. *Psychiatry, 6*(10), 409−414. doi:10.1016/j.mppsy.2007.07.004

Doyle, M., & Dolan, M. (2002). Violence risk assessment: Combining actuarial and clinical information to structure clinical judgements for the formulation and management of risk. *Journal of Psychiatric and Mental Health Nursing, 9*, 649−657. doi:10.1046/j.1365−2850.2002.00535.x

Dror, I. E. (2011). The paradox of human expertise: Why experts get it wrong. In N. Kapur (Ed.) *The paradoxical brain* (pp. 177−188). Cambridge: Cambridge University Press.

Edens, J. F., Campbell, J. S., & Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. *Law and Human Behavior, 31*(1), 53−75. doi:10.1007/s10979-006-9019-y

Falzer, P. R. (2013). Valuing structured professional judgment: Predictive validity, decision-making, and the clinical-actuarial conflict. *Behavioral Sciences and the Law, 31*, 40−54. doi:10.1002/bsl.2043

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: Systematic review and meta-analysis. *British Medical Journal, 345* (e4692). doi:10.1136/bmj.e4692

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21* (1), 1−21. doi:10.1037/a0014421

Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121−137.

Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.). *Handbook of violence risk assessment* (pp. 1−18). New York: Routledge.

Hilton, N. Z., Carter, A., Harris, G. T., & Sharpe, A. J. B. (2008). Does using nonumerical terms to describe risk aid violence risk communication? *Journal of Interpersonal Violence, 23*, 171−188. doi:10.1177/0886260507309337

Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment the empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice, 23*(4), 363−382. doi:10.1177/1043986207309436

Hurducus, C. C., Singh, J. P., de Ruiter, C., & Petrila, J. (2014). Violence risk assessment tools: A systematic review of surveys. *International Journal of Forensic Mental Health, 13* (3). doi:10.1080/14999013.2014.942932

Khiroya, R., Weaver, T., & Maden, T. (2009). Use and perceived utility of structured violence risk assessments in English medium secure forensic units. *Psychiatric Bulletin, 33*,129−132. doi:10.1192/pb.bp.108.019810

Large, M., & Nielssen, O. (2001). An audit of medico-legal reports prepared for claims of psychiatric injury following motor vehicle accidents. *Australian and New Zealand Journal of Psychiatry, 35*(4). 535−540.

McSherry, B. (2004) *Risk assessment by mental health professionals and the prevention of future violent behaviour. Trends & Issues in Crime and Criminal Justice, no. 281*. Canberra: Australian Institute of Criminology.

Monahan, J. (1981). *The clinical prediction of violent behavior*. Washington, DC: National Institute of Mental Health.

Mullen, P. E. (2007). Dangerous and severe personality disorder and in need of treatment. *The British Journal of Psychiatry, 190*(49), s3−s7. doi:10.1192/bjp.190.5.s3

Mullen, P. E., & Ogloff, J. R. P. (2009). Assessing and managing the risks of violence towards others. In M. G. Gelder, N. C. Andreasen, J. J. Lopez-Ibor Jr. & J. R. Geddes (Eds.). *New Oxford textbook of psychiatry volume 2* (2nd ed). Oxford: Oxford University Press.

Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on psychopathy checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior, 32*(4), 352−362. doi:10.1007/s10979-007-9097-5

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis) agreement on risk assessment measures in sexually violent predator proceedings. Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law, 15*(1), 19−53. doi:10.1037/a0014897

National Justice Compania Naviera SA v Prudential Assurance Co Ltd (The Ikarian Reefer), 1993. 2 Lloyd's Rep 68

Neal, T. M. S., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law, 20*(2), 200−211. doi:10.1037/a0035824

Nielssen, O. (2013). Scientific and ethical problems with risk assessment in clinical practice. *Australian & New Zealand Journal of Psychiatry, 47*(12), 1198−1199. doi:10.1177/0004867413498278

Nielssen, O., Elliott, G., & Large, M. (2010). The reliability of evidence about psychiatric diagnosis after serious crime: Part I. Agreement between experts. *Journal of the American Academy of Psychiatry and the Law Online, 38*(4), 516−523.

Pressman, D. E., & Flockton, J. (2010). *VERA-2. Violent extremism risk assessment, version 2 manual* (unpublished manuscript).

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Rossegger, A., Gerth, J., Seewald, K., Urbaniok, F., Singh, J. P., & Endrass, J. (2013). Current obstacles in replicating risk assessment findings: A systematic review of commonly used actuarial instruments. *Behavioral Sciences and the Law, 31*, 154−164. doi:10.1002/bsl.2044

Shepherd, S. M., Luebbers, S., & Dolan, M. (2013). Gender & ethnicity in juvenile risk assessment. *Criminal Justice and Behaviour 40* (4), 388−408. doi:10.1177/0093854812456776

Shepherd, S. M., Luebbers, S., Ferguson, M., Ogloff, J. R. P., & Dolan, M. (2014). The utility of the SAVRY across ethnicity in Australian young offenders. *Psychology, Public Policy and Law, 20*(1), 31−45. doi:10.1037/a0033972

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences and the Law, 31*, 8−22. doi:10.1002/bsl.2052

Singh, J. P., Desmarais, S. L., Sellers, B. G., Hylton, T., Tirotti, M., & Van Dorn, R. A. (2014). From risk assessment to risk management: Matching interventions to adolescent offenders' strengths and vulnerabilities. *Children and Youth Services Review, 47*(1), 1−9. doi:10.1016/j.childyouth.2013.09.015

Singh, J. P., Desmarais, S. L., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences and the Law, 31*, 55−73. doi:10.1002/bsl.2053

Singh, J. P., Fazel, S., Gueorguiveva, R., & Buchanan, A. (2014). Rates of violence in patients classified as high risk by structured risk assessment instruments. *The British Journal of Psychiatry, 204*, 180−187. doi:10.1192/bjp.bp.113.131938

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression

analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*(3), 499−513. doi:10.1016/j.cpr.2010.11.009

Singh, J. P., Grann, M., & Fazel, S. (2013). Authorship bias in violence risk assessment? A systematic review and meta-analysis. *PLoS ONE 8*(9), e72484. doi:10.1371/journal.pone.0072484

Szmukler, G., Everitt, B., & Leese, M. (2012). Risk assessment and receiver operating characteristic curves. *Psychological Medicine, 42*(5), 895−898. doi:10.1017/S003329171100208X

Steadman, H. J., & Cocozza, J. J. (1974). *Careers of the criminally insane: Excessive social control of deviance*. Lexington, MA: Lexington Books.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124−1131.

Ward, T., & Large, M. (2013). Little evidence for the usefulness of violence risk assessment. *The British Journal of Psychiatry, 202*, 468. doi:10.1192/bjp.202.6.468

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). Assessing risk of violence to others. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 251−275). New York: The Guilford Press.