Chapter 3
Strengths of Actuarial Risk Assessment

Robert J.B. Lehmann, Yolanda Fernandez, and Leslie-Maaike Helmus

Forensic assessment done well is a comprehensive process of obtaining information from diverse sources and creating an integrated conceptualization of the information in order to understand the client, inform decision makers, provide appropriate intervention, and manage future risk. This task is an important part of many legal decisions (e.g., civil commitment evaluations, end of sentence evaluations, and allocation of treatment) as the potential danger to society of individuals who are already known to have committed a violent offense constitutes a major concern for courts and forensic practitioners. A critical part of the process is risk assessment, which involves combining multiple risk factors together into an overall assessment of the likelihood of an outcome, such as recidivism (Hanson & Morton-Bourgon, 2009). Risk assessment and risk measures have evolved considerably over the last decades (e.g., Hanson, 2005; Harris & Hanson, 2010; Mann, Hanson, & Thornton, 2010) and distinct approaches to and generations of risk assessment can be differentiated (Andrews & Bonta, 2010; Bonta, 1996; Heilbrun, 1997).

Heilbrun (1997) argues that there are at least two models of risk assessment: the prediction and the management model. The prediction model focuses on maximizing the accuracy of the prediction of the outcome—in this model, it does not matter why something predicts the outcome, just that it does. The management model aims

R.J.B. Lehmann (✉)
Institute of Forensic Psychiatry, Charité University Medicine Berlin,
Oranienburger Straße 285, Berlin 13437, Germany
e-mail: R.Lehmann@charite.de

Y. Fernandez, C.Psych.
Correctional Service Canada\Service Correctionnel Canada, Regional Headquarters
(Ont)\Bureau Régional (Ont), Kingston, ON, Canada K7L 4Y8
e-mail: yolanda.fernandez@csc-scc.gc.ca

L.-M. Helmus
Forensic Assessment Group, 11 Aspen Grove, Nepean, Ottawa, ON, Canada
e-mail: Lmaaikehelmus@gmail.com

at reducing the risk of the occurrence of a specified event's outcome (e.g., sexual recidivism). In contrast, Bonta (1996) has provided a similar but more nuanced characterization of the development of risk assessment in three generations. The first generation consists of unstructured clinical judgment (UCJ), where a clinician gathers information and forms a subjective risk assessment. The weaknesses of this method are its overreliance on personal discretion and its lack of accountability and replicability (Bonta, 1996).

The second generation of risk assessment relies on instruments that combine primarily static (i.e., historical and unchanging), empirically derived risk factors (Bonta, 1996). In these instruments (commonly referred to as actuarial), items are often scored with either a 0–1 dichotomy (absent-present) or with a specified weighting determined by the strength of the item's relationship to recidivism. The weakness in this generation is that the focus on static factors is assumed to preclude identification of areas to target in treatment to reduce risk and it cannot reflect positive changes (Bonta, 1996).

The third generation evolved from the second to incorporate criminogenic needs (Bonta, 1996), which are dynamic (i.e., changeable) risk factors that, if changed, should alter the likelihood of reoffending (Andrews et al., 1990). Examples of key criminogenic needs (Andrews & Bonta, 2010) include antisocial personality (e.g., aggression, impulsivity) and antisocial attitudes (e.g., negative attitudes toward the criminal justice system, identification with criminals). Third-generation scales are therefore sensitive to offender changes and they also tend to have a stronger basis in theories of offending, as well as empirical evidence (Bonta, 1996). Similar to the second generation, these tools are typically actuarial. Recently, Andrews, Bonta, and Wormith (2006) have suggested that a fourth generation of risk assessment has emerged, which provides a comprehensive guide for human service delivery that spans from intake through to case closure.

In terms of understanding dynamic risk factors (i.e., third- and fourth-generation approaches), Hanson and Harris (2000) have articulated a further distinction between stable and acute dynamic factors. Stable factors constitute relatively enduring problems (e.g., alcoholism, personality disorders) and acute risk factors are rapidly changing features indicating imminent risk of reoffending (e.g., intoxication, emotional collapse). Whereas the strength of stable risk factors is monitoring risk over the medium to long term (e.g., treatment change), acute risk factors are intended for monitoring current risk over a high-risk period (e.g., community supervision).

One area not addressed by Bonta's (1996) description is the status of structured professional judgment (SPJ). SPJ is a method of risk assessment where explicit risk factors (often both static and dynamic) are scored, but the combination of these items into an overall evaluation of risk is left to the judgment of the clinician (Boer, Wilson, Gauthier, & Hart, 1997). Proponents of SPJ argue that clinical judgment should be incorporated in risk assessment because the statistical approach of actuarial scales is not always appropriate in individual cases (Webster, Douglas, Eaves, & Hart, 1997). SPJ therefore has the greatest amount of flexibility to respond to unique case-specific factors. Other researchers, however, have been dismissive of SPJ (Andrews & Bonta, 2010; Bonta, 2002; Quinsey, Harris, Rice, & Cormier,

2006) and classify it as a variation of the first generation of risk assessment (Andrews et al., 2006).

Hanson and Morton-Bourgon (2009) have added to the classification of risk assessment methods by applying a more stringent definition of actuarial scales. Their definition is based on Meehl's (1954) criteria that actuarial scales involve explicit rules to combine pre-specified items into total scores and empirically derived estimates of recidivism probability linked to each total score (Hanson & Morton-Bourgon, 2009). Given that several tools satisfying the first criteria of actuarial scales do not include absolute recidivism estimates, Hanson and Morton-Bourgon (2009) made a distinction between actuarial scales (using Meehl's definition) and mechanical scales. Mechanical scales typically contain factors identified based on theory or previous literature reviews, which are combined into a total score based on explicit item weightings, but do not contain a table with recidivism estimates per score. If SPJ scales are used to sum items to produce a total score, without creating a summary professional judgment, this would be using the SPJ scale as a mechanical scale.

The purpose of this chapter is to discuss the strengths of actuarial risk assessment. First, we will provide greater discussion of ways to conceptualize risk factors that may be included in risk scales (actuarial or other approaches). Then, we will discuss what types of information can be provided by actuarial risk scales, how the greater objectivity inherent in actuarial risk scales contributes to understanding important psychometrics of the risk assessment approaches, and how the predictive accuracy of actuarial scales compares to other approaches. These sections will be applicable to any type of offender risk assessment (i.e., any scale designed to predict an outcome among offenders). In the next section, the reader will be introduced to a small sampling of sexual offender risk scales. Sex offender risk scales are focused on because we have greater familiarity with them and they will serve as examples of the types of scales that could be used with other offender types. Then, results of surveys will be highlighted to illustrate what scales are being used in practice and how the information is being used. Lastly, the practical clinical power of actuarial risk assessment instruments in everyday practice will be discussed.

Conceptualizing Risk Factors: Psychologically Meaningful Risk Factors

As discussed above regarding the generations of risk assessment (Bonta, 1996), risk factors have often been classified as either static or dynamic (with dynamic factors further classified as stable or acute). The assumption has been that only dynamic risk factors can identify treatment targets or be used in risk management models. As an alternative to the static/dynamic conceptualization of risk factors, however, another approach is to focus on psychologically meaningful risk factors (Mann et al., 2010), also sometimes called risk-relevant propensities. In this model, risk factors are indicators of underlying constructs/propensities. For example, self-regulation problems may be an underlying psychological propensity related to recidivism.

Certain past and present behaviors, such as substance abuse, job instability, getting into fights, and poor problem-solving, may all be indicators of this propensity. In this model, the distinction between static and dynamic risk factors is simply a heuristic to describe indicators, rather than a fundamental difference between the risk-relevant constructs. For example, a history of car accidents (a static variable) and current substance abuse (a dynamic variable) may both be indicators of the same underlying propensity (poor self-regulation). In other words, psychologically meaningful risk factors can be measured using either static or dynamic risk factors.

Nonetheless, even though static and dynamic risk factors may measure the same constructs, there are practical advantages to distinguishing between them in risk assessment. Conceptually, it is easy to divide risk factors into those that the offender cannot change or manage (static) versus those he/she can (dynamic), with the latter being easier to incorporate into treatment planning (though this does not mean that static risk assessment cannot also inform risk management). Also, the types of information used to assess these risk factors are different. Static risk factors are often easy and reliably coded based on fairly straightforward criminal history information, as well as offender and victim demographics. Interviews with the offender may not be required, which makes these items practical for correctional systems that need to assess and manage large populations with limited resources. In comparison, dynamic risk factors are often more time-intensive to assess. Credible assessments should minimally include detailed reviews of file information (criminal history and personal/social history) and ideally an interview with the offender (e.g., Fernandez, Harris, Hanson, & Sparks, 2014). Other sources of information (e.g., specialized testing, collateral interviews) can also enhance dynamic assessment.

Complicating this distinction further is recent research and theoretical work that suggests the existence of *protective* factors (e.g., Farrington & Ttofi, 2011; Lösel & Farrington, 2012), which may reduce the risk of recidivism or interact with a risk factor to decrease its association with recidivism. Although the attempt to focus on offender strengths in assessment is admirable and would likely increase the comprehensiveness of the assessment and improve the therapeutic climate, Harris and Rice (2015) have argued that current descriptions of supposedly protective risk factors are mostly just the opposite end of risk factors and do not reflect new constructs. Consequently, the idea of risk-relevant propensities (Mann et al., 2010) implies that static, dynamic, and/or protective factors can be used to assess the same risk-relevant contructs, thereby informing risk management practices. Certainly, however, assessing *changes* in risk would require some consideration of dynamic risk factors.

Crime Scene Behaviors as Indicators of Risk-Relevant Propensities

One neglected area of research has been to use crime scene behaviors as indicators of risk-relevant constructs. Enduring risk-related individual offender propensities (e.g., hostility) may manifest themselves in concrete offense behavior

(e.g., excessive humiliation, genital injury). Consequently, research trying to understand offender characteristics from crime scene behavior may be relevant to risk assessment.

Canter and Heritage were among the first researchers to classify sexual offenders on the basis of observable or directly inferred crime scene behavior alone. In essence, this task consists of analyzing largely observable behaviors with inferences made about the latent (or unobservable) dimensions and themes within the data. Loosely, this process is referred to as Behavioral Thematic Analysis (BTA), a cornerstone of investigative psychology (IP) research (Canter, 2004). BTA has been used as a predictive tool exploring the relationship between behavioral themes and stranger offender characteristics with notable success (e.g., Goodwill, Alison, & Beech, 2009; Häkkänen, Puolakka, & Santtila, 2004; Mokros, 2007; Santtila, Häkkänen, Canter, & Elfgren, 2003).

Studies employing BTA of stranger rape offense details have found the presence of five (Canter & Heritage, 1990), four (Alison & Stein, 2001; Canter, Bennell, Alison, & Reddy, 2003) or three (Canter, 1994; Häkkänen, Lindlöf, & Santtila, 2004) themes of offense behavior. Although the BTA of these previous studies differed in interpretation, it is argued, in line with Wilson and Leith (2001), that each was consistent in finding themes of hostility, criminality, and pseudo-intimacy. The hostility theme is characterized by expressive, non-strategic aggression beyond that necessary to commit the offense. Here, the offender wants to hurt the victim and may perform brutal (sadistic) sexual acts. In the criminality theme, the sexual assault is considered one among many antisocial behaviors the offender commits. Whereas for stranger rapists the pseudo-intimacy theme may represent deviant sexual fantasies involving the victim receiving intense pleasure during the offense and falling in love with the offender, for the acquaintance rapist this theme may represent the misperception of the victim's sexual intent. However, during the offense both offender types show behaviors frequently present in consensual relationships.

Similarly, studies employing BTA of child molestation offenses have found the presence of three (Canter, Hughes, & Kirby, 1998) or four (Bennell, Alison, Stein, Alison, & Canter, 2001) offense themes. Here, it is argued that these themes can be summarized as fixated (i.e., love, intimate), regressed (i.e., autonomy), aggression (i.e., hostility), and criminality (i.e., control, criminal-opportunist). The themes of criminality and aggression show considerable overlap with the offense behaviors of rapists. The theme of fixation describes offenders actively creating opportunities to offend by grooming potential victims with attention, affection, and gifts and actively seeking suitable targets. The theme of regression describes offenders motivated by non-paraphilic sexual excitation and victim availability, who could choose children as an alternative to age-appropriate partners.

However, the relevance of these behavioral themes as indicators of enduring offender propensities in the context of risk assessment has been previously neglected. Therefore, based on theoretical considerations (e.g., Ward, Polaschek, & Beech, 2005) and the discussed empirical evidence (e.g., Canter et al., 2003), Lehmann and colleagues developed precise and detailed conceptualizations of target propensities and their theoretical contexts to define crime scene behavior-based indicators of

these constructs. In a first step Lehmann and colleagues were able to demonstrate the construct validity of the behavioral themes through correlational analyses with known sexual offending measures, criminal histories, offenders' motivation, and offense characteristics. For stranger rapists (Lehmann, Goodwill, Gallasch-Nemitz, Biedermann, & Dahle, 2013), the analyses revealed three behavioral offender propensities: sexuality, criminality, and hostility. Statistical analyses indicated that the behavioral theme of criminality significantly predicted sexual recidivism (AUC = 0.64) and added incrementally to Static-99. For acquaintance rapists (Lehmann, Goodwill, Hanson, & Dahle, 2015), results indicated that the behavioral themes of hostility (AUC = 0.66) and pseudo-intimacy (AUC = 0.69) predicted sexual recidivism, with the latter adding incrementally to Static-99. For child molesters (Lehmann, Goodwill, Hanson, & Dahle, 2014), the behavioral themes of fixation on child victims (AUC = 0.65) and (sexualized) aggression (AUC = 0.59) significantly predicted sexual recidivism and added incrementally to Static-99. Recently, the predictive validity of the behavioral theme of fixation was cross validated with an independent sample (Pedneault, 2014). In sum, the results indicate that crime scene information can be used to assess risk-relevant constructs. Also, crime scene information seems to be relevant external information to the results of actuarial scales.

What Types of Information Can Actuarial Risk Scales Provide?

Risk assessment can include static, dynamic, protective, or crime scene behavior factors as indicators of risk-relevant propensities. Regardless of what types of risk factors are used, how they are combined, or how accurate the scale is, appropriately reporting risk assessment results make little difference if the decision makers do not understand the information, which is a serious possibility (e.g., Varela, Boccaccini, Cuervo, Murrie, & Clark, 2014). Consequently, there have been essential developments in actuarial risk assessment research regarding optimal ways to report and interpret risk assessment information in clinical practice (for a review, see Hilton, Scurich, and Helmus, 2015). Hence, an important advantage of actuarial risk assessment instruments is that they allow their scores to be linked to different types of empirically derived quantitative indicators of risk. In contrast, other approaches to risk assessment (e.g., SPJ) solely provide nominal risk categories (e.g., low, moderate, and high risk)[1] with research indicating that nominal risk categories are interpreted inconsistently by professionals (Hilton, Carter, Harris, & Sharpe, 2008; Monahan & Silver, 2003). Three important metrics for risk communication are percentile ranks, risk ratios, and absolute recidivism rates.

---

[1] The only exception we are aware of is that the Spousal Assault Risk Assessment guide (SARA) includes percentile distributions for the total scores and number of risk factors present, although not for the overall summary judgment (Kropp & Gibas, 2010).

Percentiles

Percentiles communicate information about how common or unusual a person's score is in comparison to a reference population (Crawford & Garthwaite, 2009). Percentiles have the advantage of being fairly easily defined and communicated and are consistent with the communication of many types of psychology tests, such as intelligence tests (for more information, see Hanson, Lloyd, Helmus, & Thornton, 2012). They are particularly helpful in decisions for resource allocation. For example, if a correctional service has sufficient resources to offer treatment to 15 % of their offenders, then all the information required by an offender risk assessment may be a percentile (e.g., the highest risk 15 % should be prioritized for treatment).

Disadvantages of this metric are that the information provided is norm-referenced (i.e., relative to other offenders), when risk assessment is often intended to be criterion-referenced (i.e., focused on the likelihood of recidivism). Additionally, the relationship between percentiles and the ultimate outcome of interest (recidivism) is not necessarily linear. In other words, the difference between two risk scores in percentile units may have little to do with the difference between two risk scores in terms of the likelihood of recidivism. For example, in Static-99R, scores of −3 and −2 correspond to the 1st and 4th percentiles, respectively (with percentiles defined as a midpoint average; Hanson et al., 2012). In the higher risk range, scores of 7 and 8 correspond to the 97th and 99th percentile, respectively, which is a similar difference as scores of −3 compared to −2. In contrast, the expected recidivism rates in routine correctional samples for scores of −3 and −2 barely have a perceptible difference (0.9 % versus 1.3 %, respectively), whereas the difference in recidivism rates for scores of 7 and 8 is larger and more meaningful (27.2 % versus 35.1 %; Phenix, Helmus, & Hanson, 2015).

Risk Ratios

Risk ratios describe how an offender's risk of recidivism compares to some reference group (e.g., low risk offenders or offenders with the median risk score). For example, offenders with a Static-99R score of 4 are roughly twice as likely to sexually reoffend as offenders with a Static-99R score of 2 (Hanson, Babchishin, Helmus, & Thornton, 2013). Risk ratios are well-matched to the fundamental attribute being measured by risk scales (scorewise increases in relative risk for recidivism) and are robust to changes in recidivism rates across different samples as well as across different lengths of follow-up (Babchishin, Hanson, & Helmus, 2012a; Hanson et al., 2013). They also have the most potential for combining results from different risk scales because it is possible for them to have a common meaning across scales (Babchishin, Hanson, & Helmus, 2012b; Hanson et al., 2013; Lehmann et al., 2013).

Despite these advantages, risk ratios have rarely been developed or reported for forensic risk scales. They are, however, commonly used for medical risk communication. Possible barriers to their use include more complex calculations compared

to other metrics for communicating risk (for an example of different types of risk ratios and other decisions required in their calculation, see Hanson et al., 2013), difficulty in communicating them to laypeople (e.g., Varela et al., 2014), and potential for misinterpretation. Specifically, risk is generally overestimated if risk ratios are not properly contextualized with information about base rates (Elmore & Gigerenzer, 2005). In the Static-99R example above, knowing that an offender with a score of 4 is twice as likely to reoffend as an offender with a score of 2 has a very different meaning if the recidivism rate for a score of 2 is 4 or 40 %.

Absolute Recidivism Estimates

Absolute recidivism estimates are by far the most frequent quantitative metric reported for actuarial risk scales. They are reported in approximately 90 % of assessment reports for preventative detention in Canada, compared to percentiles and risk ratios, which are reported in roughly 40 % and 0 % of cases, respectively (Blais & Forth, 2014). In a survey examining Static-99R reporting practices in sex offender civil commitment evaluations, absolute recidivism estimates were used by 83 % of respondents, compared to roughly one third who used either percentiles or risk ratios (Chevalier, Boccaccini, Murrie, & Varela, 2014).

Absolute recidivism estimates can be generated in a variety of ways, such as from observed recidivism rates for a group of scores (ideally requiring large sample sizes for each score) or using methods such as survival analysis or logistic regression (for discussion, see Hanson, Helmus, and Thornton, 2010). Absolute risk information is easy to understand but hard to obtain with high levels of confidence. Recidivism rates vary based on the follow-up length, so this must be specified. Additionally, there are several practical complications in obtaining good estimates of recidivism, including underreporting of offences, misclassification (e.g., sexual offences pled down to nonsexual violent offences), prosecutorial discretion, and legal/policy/cultural changes over time.

Likely due to the myriad factors that influence recidivism, research has found that absolute recidivism estimates were unstable across samples for the Static-99R and Static-2002R (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012), as well as the MATS-1 (Helmus & Thornton, 2014) and the Risk Matrix 2000/Violence scale (but not the Risk Matrix 2000/Sex scale; Lehmann, Thornton, Helmus, & Hanson, 2015). Additional research has also raised concerns about the generalizability of the recidivism estimates for the VRAG (Mills, Jones, & Kroner, 2005; Snowden, Gray, Taylor, & MacCulloch, 2007). Moreover, analyses of two samples found that violent recidivism rates differed between samples after controlling for the VRS-SO pretreatment score (Olver, Beggs Christofferson, Grace, & Wong, 2014). Some solutions have been proposed for using absolute recidivism estimates in light of this variability (e.g., Hanson, Thornton, Helmus, & Babchishin, 2015), but the adequacy of these solutions is not yet known. Minimally, these findings of variability suggest that creating and reporting reliable and generalizable recidivism estimates for actuarial scales are more complicated than previously believed.

Psychometric Properties of Risk Scales

An important advantage of actuarial risk assessment is that (in contrast to UCJ) it is possible to test the psychometric properties of the risk scales. Compared to SPJ, the increased structure and availability of quantitative risk communication metrics in actuarial scales may provide more options and precision for evaluating psychometric properties, as well as stronger results. Professional standards dictate that forensic psychologists should have expertise on research related to the psychometric properties, appropriate uses, and strengths/weaknesses of risk assessment instruments they are using (American Psychological Association, 2013; Association for the Treatment of Sexual Abusers, 2014). The ability to comment on the psychometric properties of a risk scale is particularly important when risk decisions have to be defended in court; without this information, the method of risk assessment may be considered inadmissible evidence. This section discusses appropriate and inappropriate psychometric properties of actuarial risk scales and where applicable compares them to SPJ approaches.

Objectivity and Interrater Reliability

As actuarial risk assessment scales generally rely on explicitly defined predictor variables with specific scoring rules (e.g., how much weight to give the item), this facilitates more objective, transparent, standardized, and fair assessments. In contrast, UCJ has none of these features. SPJ scales may have explicitly defined predictor variables (contributing to greater objectivity than UCJ), but the subjectivity in how they influence the overall judgment should come at the expense of some objectivity, transparency, and standardization. This objectivity should increase interrater reliability, which refers to the consistency in scores across independent raters (i.e., if two different evaluators score the same individual, will they obtain the same results?). Not only does interrater reliability increase the general validity and defensibility of the assessment, but higher interrater reliability has also been associated with significantly higher predictive accuracy in some analyses (Hanson & Morton-Bourgon, 2009). Supporting the idea that the objectivity of actuarial assessment lends itself to higher interrater reliability is a finding from the Spousal Assault Risk Assessment guide (the SARA), where the interrater reliability of the SPJ summary risk rating was considerably lower than for the total score (summing the items; Kropp & Hart, 2000).

Internal Reliability

Another metric sometimes applied to risk scales is internal consistency, which refers to the degree of interrelatedness among the items (Cortina, 1993). Cronbach's α (Cronbach, 1951) is one of the most common indices of internal consistency. Unfortunately, despite its frequent use, internal consistency is not an informative metric for actuarial risk scales.

Developing a scale to predict an outcome (e.g., recidivism) is meaningfully different than classical scale construction in psychology. Specifically, most scales in psychology are norm-referenced, which means they are trying to capture how individuals display different amounts of some relevant construct (e.g., Aiken, 1985). Examples include tests of intelligence, ability, or personality. In contrast, risk assessment scales are inherently criterion-referenced, which means they are designed specifically to predict an outcome of interest. This means that some elements of test reliability and validity are not applicable (e.g., internal consistency; Aiken, 1985). In norm-referenced scales, internal reliability increases to the extent that multiple items are assessing the same construct (e.g., items are highly related to total scores); this may be achieved by including similar items but with different wordings or reverse-scored.

In contrast, the most important goal of criterion-referenced scales is to predict the outcome. For that reason, it does not make sense (and may be undesirable) to measure only one construct and to include multiple items assessing the same construct. Consequently, predictive accuracy and efficiency are maximized by including the smallest number of items measuring the most distinct constructs possible, instead of having multiple items assess a single construct. These goals would deliberately decrease internal consistency. Consequently, we do not recommend reporting internal consistency to evaluate the reliability of risk scales. Internal consistency is, however, useful for scales designed to assess a single construct (e.g., the Psychopathy Checklist-Revised; Hare, 2003).

Construct Validity

The results of risk scales should have greater meaning and clearer implications for case management decisions when the source of an offender's risk is identified and understood. This requires knowing what constructs are being measured by actuarial risk scales. Given that risk scales were designed as criterion-referenced (i.e., items were chosen based on their ability to predict the outcome), construct validity has been largely neglected in actuarial risk assessment scales. In recent years, however, greater attention has been paid to construct validity of actuarial risk scales (e.g., Babchishin et al., 2012b; Brouillette-Alarie, Babchishin, Hanson, & Helmus, 2015).

Specifically, items are assumed to predict the outcome because they are an indicator of some kind of latent underlying construct/propensity (Mann et al., 2010). Efforts to improve construct validity may focus on identifying the underlying constructs measured by the items, determining how well the items measure those constructs, and assessing how to best combine constructs into an overall assessment. Consequently, greater focus on construct validity should help improve predictive accuracy (by potentially identifying better indicators of constructs), resolve discrepancies in risk scales, identify optimal ways to combine risk scales, and better identify whether external information is likely to add to the results of an actuarial scale (e.g., Hanson, 2009).

Predictive Validity

Whereas reliability specifies the extent to which risk assessments give consistent results, predictive validity refers to the accuracy of measurement in predicting the outcome. For risk assessment, predictive validity (also called criterion-related validity) is most important. Discrimination and calibration are distinct indices of the predictive validity of a criterion-referenced scale (Altman, Vergouwe, Royston, & Moons, 2009).

Discrimination quantifies the model's ability to distinguish between recidivists and non-recidivists or in other words, to rank offenders according to their relative risk to reoffend. This indicates whether higher risk offenders are more likely to reoffend than lower risk offenders. The most commonly recommended and reported statistic for discrimination is the area under the curve from receiver operating characteristic curve analyses (AUC; Mossman, 1994; Swets, Dawes, & Monahan, 2000). For further discussion of the strengths and weaknesses of other discrimination statistics (such as correlations, Harrell's *c* index, and Cox and logistic regression), see Babchishin and Helmus (2015).

In contrast, there is little research on the calibration of risk scales, which refers to the ability of a risk scale to estimate absolute recidivism rates (Helmus, Hanson et al. 2012). Consequently, there are no well-established statistics for measuring calibration. For example, in 2009 there were at least 63 studies examining the discrimination of Static-99 (summarized in Hanson & Morton-Bourgon, 2009) but only two studies that examined its calibration (Doren, 2004; Harris et al., 2003). One promising statistic to assess calibration is the E/O index (Gail & Pfeiffer, 2005; Rockhill, Byrne, Rosner, Louie, & Colditz, 2003), which is the ratio of the predicted number of recidivists (E) divided by the observed (O) number of recidivists (Viallon, Ragusa, Clavel-Chapelon, & Bénichou, 2009; for more discussion of this statistic, see Helmus and Babchishin, 2014). Although calibration statistics have been historically neglected, they present one of the most promising advantages of actuarial risk scales. Discrimination can be examined with either SPJ or actuarial approaches, but calibration is a unique property of actuarial risk scales, as they are the only approach with empirically derived recidivism estimates associated with total scores.

Predictive Accuracy of Actuarial Scales Compared to Other Approaches

Research across a variety of disciplines (including offender risk assessment) supports the superiority of actuarial prediction schemes over professional judgment (Ægisdóttir et al., 2006; Bonta, Law, & Hanson, 1998; Dawes, Faust, & Meehl, 1989; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009; Mossman, 1994). Examining sex offender risk assessment, for example, recent meta-analytic research (Hanson & Morton-Bourgon, 2009), has found that actuarial measures had significantly higher accuracy in predicting sexual recidivism

($d = 0.67$) than UCJ ($d = 0.42$), whereas SPJ scales had accuracy closer to UCJ, but not significantly different than either of the two previous categories ($d = 0.46$).

This cross-disciplinary literature contradicts the intuitive belief that the expertise of professionals should be better equipped to handle complex situations and case-specific factors (e.g., Boer et al., 1997). Paradoxically, it appears to be simultaneously correct that although level of expertise matters (e.g., experts generally outperform novices), actuarial decision algorithms outperform experts, but only under some conditions (Kahneman & Klein, 2009; Shanteau, 1992). An important question, then, is under what conditions?

In summarizing decision-making and cognitive science literature, Shanteau (1992) found evidence for good expert performance in weather forecasters, livestock judges, astronomers, test pilots, soil judges, chess masters, physicists, mathematicians, accountants, grain inspectors, photo interpreters, and insurance analysts. Poor professional judgments were noted for clinical psychologists, psychiatrists, astrologers, student admissions evaluators, court judges, behavioral researchers, counselors, personnel selectors, parole officers, polygraph judges, intelligence analysts, and stock brokers. Mixed performance was found for nurses, physicians, and auditors. Shanteau (1992) proposed a variety of task features that were associated with poorer performance from experts. He concluded that human behavior is inherently more unpredictable than physical phenomena and that decision-making is particularly difficult for unique tasks, when feedback is unavailable and when the environment is intolerant of error.

Kahneman (2011) provided a more updated summary of the performance of experts across a variety of tasks, with similar conclusions. According to Kahneman and Klein (2009), expert opinion can be expected to outperform actuarial decisions when the environment is regular (i.e., highly predictable), the expert has considerable practice, and there are opportunities to get timely feedback on decisions to learn from errors or false cues. These conditions are generally not present in offender risk assessment. The sheer number of diverse predictors of recidivism (e.g., see Andrews and Bonta, 2010, and Hanson and Morton-Bourgon, 2005) suggests that criminal behavior is not highly predictable (i.e., the number of contingencies are infinite; Hanson, 2009), and evaluators do not receive timely feedback on their decisions.

Professional Overrides

Another way to compare the predictive accuracy of actuarial approaches to SPJ is to examine "professional overrides." A professional override is when the results of an actuarial scale are adjusted based on professional judgment. The premise of SPJ scales is that the professional judgment is a helpful way to respond to case-specific factors or apply flexibility in terms of weighting items for a particular individual. Research, however, has consistently found that overrides to actuarial scales decrease their accuracy (Hanson, Helmus, & Harris, 2015; Hanson & Morton-Bourgon, 2009; Wormith, Hogg, & Guzzo, 2012). Research also demonstrates that professional

judgment tends to be more conservative, less transparent, and less replicable than are actuarial measures (Bonta & Motiuk, 1990). Alexander and Austin (1992) have found that overrides also disproportionately are used to increase offenders' risk. If overrides are a necessary part of correctional policy (e.g., to introduce flexibility), Austin, Johnson, and Weitzer (2005) encourage adopting a general standard where only 5–15 % of final assessments should differ from initial actuarial results. Furthermore, the direction of inconsistencies should be balanced, where half are higher and half are lower than the original actuarial result. Overall, however, overrides may offer some advantages (e.g., flexibility), but the research seems clear that they have a negative impact on accuracy. One possible explanation for the disappointing results of professional judgment in this context is that the professionals may be able to accurately identify risk-relevant information that is not incorporated in the risk scale, but are unable to determine to what extent this new information is correlated with existing information in the scale or how much weight to give this new information.

Incremental Validity

Besides predictive accuracy, incremental validity which assesses the contribution of an additional measure to the prediction of an outcome (e.g., recidivism) is essential information in the context of risk assessment. Additional measures may add incrementally by either improving the measurement of constructs already included (e.g., attitudes, emotional regulation, intimacy deficits) or by the assessment of new risk-related constructs. The greater objectivity and structure of actuarial risk scales may facilitate easier interpretation of incremental results.

Incremental validity becomes increasingly important as the knowledge base for offender risk assessment expands. As risk scales become entrenched in practice, the threshold for newly developed scales should increase. In other words, if scales are already in use, the onus is on developers of new scales to demonstrate that their scale provides incremental accuracy to standard practice (Hunsley & Meyer, 2003). Unfortunately, statistical power is reduced for tests of incremental validity compared to bivariate predictive validity, and comparisons of scales may require sample sizes in the thousands (Babchishin et al., 2012b). This means that increasingly larger amounts of data are required for smaller gains in accuracy.

Combining Actuarial Risk Instruments

Generally, a comprehensive actuarial risk assessment of a range of psychological risk factors will yield better predictive accuracy than a less comprehensive assessment (Hanson & Morton-Bourgon, 2009; Mann et al., 2010). Accordingly, multiple risk measures are frequently used to assess offenders' risk for future offending (Jackson & Hess, 2007; Neal & Grisso, 2014). The use of multiple risk tools is

justified on the grounds that they provide incremental information (Babchishin et al., 2012b; Welsh, Schmidt, McKinnon, Chattha, & Meyers, 2008). For some scales the developers propose starting with a commonly used risk scale and adjusting the overall rating based on the scores of an incrementally valid, additional risk instrument (e.g., Helmus, Hanson, Babchishin, & Thornton, 2014). Also, recent research indicates that averaging the risk ratios of different risk tools is a promising approach to obtaining a better overall evaluation of relative risk (Lehmann, Hanson et al. 2013), as opposed to other approaches, such as taking the highest or lowest risk estimate. Hence, a strength of actuarial risk assessment is the inclusion of a range of empirically validated risk factors or scales, which under certain circumstances (see Lehmann, Hanson et al. 2013) could be combined into an overall risk judgment of recidivism risk with better predictive accuracy than a single scale.

Selected Examples of Actuarial Risk Scales for Sex Offenders

Below, specific examples of risk scales for sex offenders will be discussed. Note that they are not meant as an exhaustive list of scales available—they are illustrative examples of scales we are most familiar with. This chapter was not intended to provide a detailed review of actuarial risk scales available.

The Static-99/R

The most commonly used static sex offender risk assessment tools in Canada and the United States are the Static-99 and Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, Hanson, & Babchishin, 2012; Interstate Commission for Adult Offender Supervision, 2007; Jackson & Hess, 2007; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010; Neal & Grisso, 2014). The Static-99/R is 10-item actuarial scales designed to assess sexual recidivism risk of adult male sex offenders. The items and scoring rules for Static-99 (Hanson & Thornton, 2000) and Static-99R (Helmus, Thornton et al., 2012) are identical with the exception of updated age weights for the Static-99R. The scale developers have recommended that Static-99R be used in place of the original scale (Helmus, Thornton et al., 2012). Static-99/R contains items covering the broad constructs of age and relationship status (i.e., whether the offender has ever lived with a lover for two or more years), sexual deviance (e.g., stranger victims, noncontact sexual offences, prior sex offenses), and general criminality (e.g., number of prior sentencing occasions, index nonsexual violence, prior nonsexual violence) identified in meta-analytic research (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005).

Accordingly, the strength of the risk tool is that it only uses risk factors empirically associated with sexual recidivism. Also, explicit rules for combining the factors into a total risk score are provided (A. Harris, Phenix, Hanson, & Thornton, 2003).

Other advantages are that with appropriate training, the scale can be scored quickly based on commonly available demographic and criminal history information, without a detailed file review or interview with the offender. The website for the scale (www.static99.org) contains an evaluator workbook that includes normative data for interpreting Static-99/R (nominal risk categories, absolute recidivism estimates, percentiles, and risk ratios) and sample reporting templates and is regularly updated with more recent research and normative data for the scale. Although Static-99R was designed to predict sexual recidivism, normative data for violent recidivism risk has previously been available for the scale as well. Most recently, Babchishin, Hanson, and Blais (2015) have found that the inclusion of so many items assessing sexual deviance overly dilutes the scale's predictive accuracy for violent recidivism. Consequently, the developers of Static-99R no longer recommend its use to comment on violent recidivism risk among sex offenders. Instead, they recommend using the BARR-2002R (Brief Assessment of Recidivism Risk-2002R), which was created form a subset of Static-2002R items (see Babchishin et al., 2015).

In terms of the psychometric properties of the risk scale, recent meta-analyses found moderate accuracy in predicting sexual recidivism for both Static-99 ($d=0.67$, $k=63$, $n=20,010$; Hanson & Morton-Bourgon, 2009) and Static-99R ($d=0.76$, $k=23$, $n=8106$; Helmus, Hanson et al. 2012). The interrater reliability for Static-99/R reported across different samples was found to be generally high (ICC>0.75; see Anderson & Hanson, 2010; Phenix & Epperson, 2015; Quesada, Calkins, & Jeglic, 2013). Risk ratios for Static-99R have been found to be highly stable across diverse samples and time period (Hanson et al., 2013), although the absolute recidivism rates per Static-99R score have significantly varied across samples (Helmus, Hanson et al. 2012), which complicates interpretation of the scale. Current recommendations for using Static-99R in light of this base rate variability are discussed by Hanson et al. (2015).

Risk Matrix 2000

The Risk Matrix 2000 (RM2000) has been adopted by the police, probation, and prison services of England, Wales, Scotland, and Northern Ireland (National Policing Improvement Agency, 2010; Social Work Inspection, HM Inspectorate of Constabulary for Scotland, & HM Inspectorate of Prisons, 2009). The RM2000 is an actuarial scale that assesses recidivism risk of adult male sexual offenders (Thornton et al., 2003). The scale is based on file information only and contains three separate sales: one for measuring risk of sexual recidivism (RM2000/S), one for measuring risk of nonsexual violent recidivism (RM2000/V), and one combination of the first two scales for measuring risk of any violent recidivism (RM2000/C).

The scoring of the RM2000/S includes two steps. In step 1 three risk items are scored (number of previous sexual appearances, number of criminal appearances, and age at next opportunity to offend) and offenders are assigned to four preliminary risk categories. In the second step four aggravating risk factors

(any conviction for sexual offense against a male, any conviction for a sexual offense against a stranger, any conviction for a noncontact sex offense, and single – never been married) need to be considered. The presence of two or four aggravating factors raises the risk category by one or two levels, respectively. For the RM2000/V three items need to be scored (age on release, violent appearances, and any conviction for burglary) and offenders are also assigned to the four risk categories. The four nominal risk categories are low, medium, high, and very high risk. To get the score for the RM2000/C scale, the risk category points for the RM2000/S and RM2000/V need to be summed and converted into the four nominal risk categories.

In terms of the psychometric properties of the three risk scales, a recent a meta-analysis (Helmus, Babchishin, & Hanson, 2013) found moderate to high accuracy in predicting sexual recidivism for the RM2000/S (mean $d=0.74$ in both fixed-effect and random-effects models, $k=15$, $n=10,644$), in predicting nonsexual violent recidivism for the RM2000/V (after adjusting the largest study weight, mean fixed-effect $d=0.98$ and random-effects $d=0.96$, $k=10$, $n=9836$), and in predicting any violent recidivism for the RM2000/C (fixed-effect $d=0.81$ and random-effects $d=0.80$, $k=8$, $n=8277$).

Recently, Lehmann, Thornton et al. (2015) developed non-arbitrary metrics for risk communication for the RM2000 (i.e., percentiles, risk ratios, and absolute recidivism estimates) based on combining offenders from four samples of fairly routine (i.e., complete/unselected) settings: England and Wales, Scotland, Berlin (Germany), and Canada ($n=3144$). Although there were meaningful differences across these samples in the distribution of Risk Matrix scores, relative increases in predictive accuracy for each ascending risk category were remarkably consistent across samples. However, recidivism rates for the median risk category also showed some variability across samples for the Risk Matrix 2000 Violence and Combined scales, but not for the Sex scale (Lehmann, Thornton et al., 2015).

The Crime Scene Behavior Risk Measure

Whereas previous actuarial risk assessment instruments of static risk factors focused on the criminal history of sexual offenders, recent research indicates that sexual offender risk assessment can be improved by also utilizing crime scene behavior as indicators of risk for sexual recidivism. The seven items (explicit offense planning, sexualized language, actively seeking victim, no multiple juvenile offenders, approach-explicit, male victim at index offense, and hands-off: victim active) that comprise the Crime Scene Behavior Risk measure (CBR; Dahle, Biedermann, Lehmann, & Gallasch-Nemitz, 2014) showed high predictive accuracy for sexual recidivism with little variation between the development (c index[2]$=0.72$; $n=995$) and the replication sample (c index$=0.74$; $n=77$).

---

[2] Harrell's c index is an effect size analogous to the AUC, but it takes into account varying follow-up times. The c value can be interpreted in the same way as an AUC, with values of 0.56, 0.64, and 0.71 noting small, moderate, and large effect sizes.

The interrater reliability for the CBR total score ranged from moderate (ICC=0.60) in the development sample to excellent (ICC=0.89) in the cross-validation sample. For risk communication the authors provide estimated recidivism rates for each CBR score after 5 and 10 years. Further, the CBR was found to provide significant incremental validity and to improve the predictive accuracy of the Static-99R risk assessment tool (Dahle et al., 2014). Accordingly, the authors of the CBR recommend using the published nominal risk categories of the Static-99R (Helmus, Thornton et al., 2012) as an initial assessment of recidivism risk and adjusting the risk level according to the CBR score to obtain a better overall evaluation of recidivism risk. Hence, the assessment of sexual recidivism risk using different sources of information should yield a better understanding of the recidivism risk that emanates from a specific offender.

Stable-2007

The Stable-2007 (Hanson, Harris, Scott, & Helmus, 2007) is an interview- and file-review-based instrument designed to assess stable (i.e., medium- to long-term) dynamic risk factors for sexual recidivism, which are unlikely to change without deliberate effort (i.e., treatment targets; Hanson & Harris, 2013). Items are scored on a 3-point scale ranging from "0, no problem;" "1, maybe/some," to "2, yes, definite problem." The instrument contains 13 items divided into the 5 subsections of significant social influences, intimacy deficits (i.e., capacity for relationship stability, emotional identification with children, hostility toward women, general social rejection/loneliness, and lack of concern for others), sexual self-regulation (i.e., sex drive/preoccupation, sex as coping, and deviant sexual interests), general self-regulation (i.e., impulsive acts, poor cognitive problem-solving, and negative emotionality/hostility), and cooperation with supervision. The total score is obtained by summing all items and can range from 0 to 26 for offenders with a child victim and 0 to 24 for other offender types (the item *emotional identification with children* is scored only for offenders with a child victim). The Stable-2007 can inform decisions about treatment targets as well as about moderate- to long-term recidivism potential with higher scores indicating greater risk of sexual recidivism. In addition to detailed coding rules for each item, the Stable-2007 scoring manual also includes sample interview questions, practice cases, reporting suggestions, and advice for maintaining high quality risk assessments (Fernandez et al., 2014).

Excellent interrater reliability has been found for the Stable-2007 total score (ICC>0.75; Fernandez, 2008; Hanson et al., 2007). The predictive accuracy of the Stable-2007 for sexual recidivism was found to range from moderate (e.g., AUC=0.67; Hanson et al., 2015) to high (e.g., AUC=0.71; Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012).

For risk communication the authors provide nominal risk categories for the Stable-2007 as follows: 0–3=low need, 4–11=moderate need, and 12 or greater=high need, as well as percentiles (Fernandez et al., 2014). Hanson et al. (in

press) found the Stable-2007 to add incrementally to the Static-99R and Static-2002R in most analyses. Of the scales, however, the Static-99R and Static-2002R had higher predictive accuracy than the Stable-2007. Consequently, the scale developers recommend using it in conjunction with a static scale (Hanson, Helmus, & Harris, 2015). The current evaluator workbook of Stable-2007 contains 1-year, 3-year, and 5-year recidivism estimates for risk categories based on combining Stable-2007 with either Static-99R, Static-2002R, or the Risk Matrix-2000 (Helmus et al., 2014; Helmus & Hanson, 2013).

Acute-2007

The Acute-2007 (Hanson et al., 2007) is an interview- and file-review-based instrument designed to assess acute dynamic (i.e., rapidly changing) risk factors for sexual recidivism which is essential to managing sexual offenders on community supervision. Items are scored on a 4-point scale ranging from "0, no problem;" "1, maybe/some;" "2, yes, definite problem;" to "3, intervene now." The Acute-2007 includes seven items (access to victims, sexual preoccupation, hostility, rejection of supervision, emotional collapse, collapse of social supports, and substance abuse), all of which are predictive of general recidivism. For predicting sexual or violent recidivism, however, a subscale of only four items is included (the first four listed above; Hanson et al., 2007). Some subsequent analyses have suggested that the four items of the sexual/violence subscale represent more of an approach trajectory toward offending, whereas the three additional items are more indicative of an emotional collapse/avoidant trajectory toward offending (Babchishin, 2013). Scores for the sex/violence subscale can range from 0 to 12, whereas the total of the general recidivism scale can range from 0 to 21, with higher scores indicating a higher likelihood of recidivism. The cut scores for the sex/violence subscale are 0=low, 1=moderate, and 2+=high imminent recidivism risk. For the general recidivism scale, the recommended cut scores are reported as 0=low, 1–2=moderate, and 3+=high.

In the development study the interrater agreement for the individual Acute items ranged from good to excellent with a median ICC of 0.90. Feedback from users suggested that the brevity of the item descriptions in the coding manual might be contributing to subjective variability in scoring some items. Consequently, a new manual with more comprehensive item descriptions along with examples for item scoring is in development. Both the general scale (AUC=0.72) and the sex/violence subscale (AUC=0.74) showed high ability to differentiate between the imminent sexual recidivists and the non-recidivists in the development sample (Hanson et al., 2007), though the three extra items of the general scale did not predict sexual recidivism. The sex/violence subscale significantly predicted imminent (within 45 days) sexual, violent, and any recidivism after controlling for the combined Static-99/Stable-2007 categories whereas the general recidivism Acute score only added incrementally to the prediction of violent and general recidivism. Accordingly,

specific rules on how to combine static, stable, and acute factors into three priority levels were constructed by the authors. For risk communication relative risk ratios for sexual recidivism within 45 days based on combined Static-99, Stable-2007, and Acute-2007 scores are presented for the three priority levels. Recently, Babchishin (2013) investigated the temporal stability of the factor structure of the Acute-2007 and found observed changes to be attributed to true changes on risk-relevant propensities assessed by the Acute-2007, as opposed to measurement error.

Violence Risk Scale-Sexual Offender Version (VRS-SO)

The VRS-SO (Wong, Olver, Nicholaichuk, & Gordon, 2003) is a 24-item interview- and file-review-based instrument comprised of 7 static (e.g., age at release, prior sex offenses, unrelated victim) and 17 dynamic items which are scored on a 4-point Likert-type scale ranging from 0 to 3, with higher scores indicating increased risk for sexual recidivism. Factor analysis of the dynamic items generated three factors labeled sexual deviance ($\alpha=0.87$; e.g., deviant sexual preference, offense planning, sexual compulsivity), criminality ($\alpha=0.79$; e.g., impulsivity, substance abuse, compliance with community supervision), and treatment responsivity ($\alpha=0.72$; e.g., insight, treatment compliance, cognitive distortions). Accordingly, the first two factors are consistent with the two major constructs related to sexual reoffending discussed above. All 24 items are used to assess recidivism risk. However, the VRS-SO was designed to integrate sex offender risk assessment and risk reduction through treatment. Therefore, the dynamic items are used to identify treatment targets and to measure change. Here, change is measured on the basis of a modified application of the key transtheoretical constructs of stages of change (SOC; Prochaska, DiClemente, & Norcross, 1992). The progression in the SOC is supposed to indicate the extent to which the offender has improved (i.e., changed). Therefore, treatment targets (i.e., dynamic items rated 2 or 3) are given a SOC rating at pre- and posttreatment and both ratings are compared to quantify change (Olver, Wong, Nicholaichuk, & Gordon, 2007).

The developers of the scale investigated the psychometric properties of the VRS-SO (Olver et al., 2007). Excellent interrater reliability has been found for the pretreatment (ICC=0.74) and posttreatment (ICC=0.79) dynamic item total score. The predictive accuracy of the VRS-SO total score was found to be high for sexual recidivism for both pretreatment (AUC=0.71) and posttreatment (AUC=0.72). Also, both the VRS-SO static and dynamic item total scores made unique contributions to the prediction of sexual recidivism after controlling for Static-99. These findings were replicated in an independent validation study (Beggs & Grace, 2010). One limitation of this research is that similar to other scales, Olver, Beggs Christofferson, and Wong (2015) have found significant variability in the recidivism rates of two samples, even after controlling for the VRS-SO pretreatment score. Such variability poses a challenge for the creation of generalizable recidivism estimates.

Importantly, therapeutic change (i.e., positive change in dynamic items) was found to be significantly related to reduction in sexual recidivism after controlling for risk and follow-up time (Beggs & Grace, 2011; Olver et al., 2007, 2014). In their most recent risk communication efforts, Olver et al. (2015) applied an intuitively useful method of conceptualizing and communicating change to the VRS-SO. Olver and colleagues used the Clinically Significant Change model, which incorporates offenders' change relative to external standards of what is "functional" and takes into account whether the change is reliable (i.e., likely accounted for by more than measurement error). Using this technique, the authors found that Clinically Significant Change provided some unique information in predicting recidivism beyond pretreatment risk scores, and they offered examples of how this approach can facilitate risk communication.

Survey Findings: What Is Used in Applied Practice?

Several surveys have been conducted to assess practical applications of risk assessment (e.g., what scales are used and how the information is incorporated). Examining 111 risk assessment reports for preventative detention hearings in Canada (intended for offenders at high risk of violent recidivism), Blais and Forth (2014) found that over 90 % of experts (appointed by either the prosecution or appointed by the court) used an actuarial risk assessment scale, compared to 53 % who used an SPJ scale. The PCL-R (Psychopathy Checklist-Revised), designed to assess the construct of psychopathy (not as a risk assessment scale), was used in over 95 % of risk assessment reports. In terms of scales designed to assess risk of recidivism, the most commonly used scale was the Static-99, used in over 60 % of cases, which is surprising given that not all candidates for preventative detention are sex offenders. The next most commonly used scales were the VRAG (Violence Risk Appraisal Guide; 48 % of reports) and the SORAG (Sex Offender Risk Appraisal Guide; 42 % of reports), both of which are actuarial. Other risk scales were used in one quarter or less of cases.

In a particularly large study, Singh and colleagues (2014) surveyed 2135 mental health professionals who had conducted at least one violence risk assessment. Half of the respondents were from Europe, followed by 21 % from North America, 5 % from Australasia, and 3 % each from South America and Asia. Among this diverse sample, over 400 different instruments were reported as being used for violence risk assessment, although roughly half had been developed specifically for personal or institutional use only. Among the 12 most frequently used risk scales, half were actuarial and half were SPJ, with the HCR-20 (Historical Clinical Risk Management 20, an SPJ scale) reported as the most commonly used, followed by the PCL-R.

Neal and Grisso (2014) surveyed 434 psychologist and psychiatrist members of various professional associations, mostly from the United States, Canada, Europe, Australia, and New Zealand, who described 868 cases they had completed. The most common types of referrals these professionals dealt with included competence

to stand trial, violence risk, sex offender risk, insanity, sentencing, disability, child custody, civil commitment, child protection, and civil tort. Use of structured assessment tools (e.g., note this is broader than risk assessment tools and could include personality assessments) varied based on the type of assessment being conducted, with the lowest rates of structured tool use reported for competence to stand trial cases (58 %), disability cases (66 %), and civil tort cases (67 %). Sex offender risk cases were most likely to use structured tools (97 %), followed by child protection cases (93 %) and violence risk cases (89 %).

Among sex offender risk cases, Neal and Grisso (2014) found that the most frequently used tools were by far the Static-99/R or Static-2002/R (which were clumped together), used in 66 % of cases. The next most commonly used tools were all either designed to assess a single construct or were personality assessments—none were designed for sex offender risk assessments. These included the PCL-R (35 % of cases), Minnesota Multiphasic Personality Inventory (MMPI; 27 % of cases), Personality Assessment Inventory (PAI; 23 % of cases), and the Millon Clinical Multiaxial Inventory (MCMI; 17 % of cases). Other sex or violent risk assessment scales, such as the Sexual Violence Risk-20 (SVR-20), Risk for Sexual Violence Protocol (RSVP), Stable-2007, SORAG, and VRAG, were used in less than 15 % of cases. Note that the SVR-20 and RSVP are SPJ scales, whereas the others are actuarial. Similar results were found in a survey of American psychologists, conducted by Archer, Buffington-Vollum, Stredny, and Handel (2006). For adult sex offender risk assessments, Static-99 was still the most commonly used scale (mentioned by roughly half of participants), but with a smaller margin over other frequently used scales, which included the SVR-20, Minnesota Sex Offender Screening Tool-Revised (MnSOST-R), Rapid Risk Assessment for Sex Offense Recidivism (RRASOR), and the SORAG. Note that Stable-2007 did not exist when this survey was completed. These findings mirror survey results of sex offender civil commitment evaluators (Jackson & Hess, 2007) and sex offender treatment programs (McGrath et al., 2010) which found Static-99 to be the most commonly used risk scale, by a wide margin. Additionally, among treatment programs, dynamic risk scales were being more widely adopted, with Stable-2007 being the most frequently used (McGrath et al., 2010).

Other important findings from surveys pertain to how experts use information from risk scales. In SPJ scales, the only information available is a nominal risk category (with the exception of the SARA, which provides some percentile information, although not for the final risk judgment; Kropp & Gibas, 2010). In actuarial scales, it is possible to report absolute recidivism estimates. Additionally, some scales may also provide information on percentiles or nominal risk ratios. In their study of Canadian preventative detention hearings, Blais and Forth (2014) found that over 95 % of risk assessment reports mentioned a nominal risk level. For actuarial scales, roughly two thirds of reports mentioned a total score, 37 % reported a percentile, and over 90 % reported absolute recidivism estimates. For SPJ scales, although the intent of the scales is NOT to sum the risk factors, 24 % of reports also included a mechanical total score from the scale. In a more recent survey of 109 experts who use the Static-99R in Sexually Violent Predator evaluations in the

United States (Chevalier et al., 2014), 83 % included nominal risk categories and absolute recidivism in their reports, while 35 % included percentiles and 33 % include risk ratios. When asked to rank the importance of the various risk communication metrics, 54 % of the evaluators reported that absolute recidivism estimates provided the most important information about recidivism risk, compared to 25 % who felt the nominal risk categories provided the most important information.

Clinical Advantages to Actuarial Risk Assessment

Psychologists have been instrumental for more than a century in developing, validating, refining, and implementing scientifically rigorous procedures that have advanced our understanding of psychological constructs and our prediction of future behavior. Evidence-based practice, or the practice of providing services that have empirically demonstrated effectiveness for each client's needs, has become the standard among clinicians and within most organizations and has extended into the field of assessment. Hunsley and Mash (2010) note that evidence-based assessment "relies on research and theory to guide the selection of constructs to be assessed for a specific assessment purpose, the methods and measures to be used in the assessment, and the manner in which the assessment process unfolds" (p. 7). In the area of correctional intervention, the use of evidence-based assessment tools such as actuarial risk measures is the first step in a comprehensive evidence-based approach, which includes assessing the client, formulating a case conceptualization, determining the client's needs, deciding on and implementing a program of treatment, and monitoring and evaluating the outcome.

Evidenced Based Practice in Correctional Settings

There is extensive research into the basic principles that should be adhered to for human services to have the greatest positive impact. Within correctional work, research supports that the more risk, need, and responsivity factors a program adheres to, the more effective it is in reducing recidivism, while programs that do not incorporate these principles potentially increase recidivism (Dowden & Andrews, 2004; Flores, Russell, Latessa, & Travis, 2005; Lowenkamp, Pealer, Smith, & Latessa, 2006; Smith & Schweitzer, 2012; Wormith, Althouse, Reitzel, Fagan, & Morgan, 2007). Specifically, intervention is most effective when targeted proportionally to offender risk (risk principle), focusing on criminogenic needs (need principle), and matched to the learning style and needs of the offenders (responsivity principle).

Consequently, evidence-based assessment is a critical first component to an effective correctional intervention (i.e., identification of the first two principles: risk and need). As part of that approach, risk assessment tools can "facilitate decisions

about the intensity of intervention in accordance with risk needs responsivity (RNR) principles" (Hilton, 2014, p. 88), thus maximizing intervention effectiveness. However, Andrews and Dowden (2005) note that inconsistencies or a lack of implementation integrity across providers is related to differences in program outcomes. Risk assessment tools, like any part of an evidence-based intervention, must be implemented with integrity to be maximally effective. For example, two field studies examining the real-world utility of Static-99 show remarkable variability. In Texas, Static-99 demonstrated minimal accuracy in predicting sexual recidivism (AUC = 0.57; Boccaccini, Murrie, Caperton, & Hawes, 2009). In contrast, California implemented the scale with rigorous training, mentoring, and ongoing quality control policies (e.g., mandatory re-certification by users) and reported exceptionally high predictive accuracy (AUC = 0.82; Hanson, Lunetta, Phenix, Neeley, & Epperson, 2014). The discrepancy in these results from two American jurisdictions highlights the importance of implementation integrity. Additionally, Hanson et al. (2014) found meaningfully higher predictive accuracy for actuarial risk scales scored by front-line staff who were more committed to the project (defined as those who completed all the requested information). For additional suggestions on best practices for quality control, see Fernandez and colleagues (2014).

In the second half of this chapter, we argue that actuarial measures form a critical part of evidence-based practice and particularly enhance program integrity by providing a standardized and structured approach to the critical first steps (assessment) of any correctional intervention. We focus on the advantages of actuarial measures as part of implementing an effective evidence-based intervention program within a clinical practice, forensic setting, or organization. Adapting Bernfeld, Blase, and Fixsen (1990) "multilevel systems perspective," the strengths and usefulness of actuarial risk assessment instruments in clinical practice are discussed across the four levels important to human service delivery: namely, the client, program, organizational, and societal levels.

The Client Level

Actuarial risk assessment has the potential for several direct advantages for the client including providing opportunities for a collaborative working relationship with the assessor, an introduction to the therapeutic relationship and to the concept of risk, identification of treatment targets, and making the best match between the client and the appropriate type of treatment. Shingler and Mann (2006) note that risk assessment offers a unique collaborative opportunity to build rapport and set the stage for subsequent intervention. The first step of their sexual offender intervention program, the Structured Assessment of Risk and Need (SARN; Webster et al., 2006), specifically integrates collaboration into the risk assessment process. Their in-house training encourages assessors to approach the risk assessment as a critical first step in the treatment process and emphasizes that the experience the offenders have during a risk assessment can heavily impact their desire to engage in treatment and the

offenders' trust of the process. Offenders themselves have expressed the importance of contributing to the assessment, and getting their side represented, in their sense of fairness and confidence in the outcome of the risk assessment process (Attrill & Liell, 2007). A thorough assessment at the front end of treatment using measures that identify factors empirically related to recidivism can help to focus the client on the important issues necessary for offenders to be able to identify and cope with risk factors to reduce the risk of recidivism (Proulx, Tardiff, Lamoureeux, & Lussier, 2000), saving them both time and effort as they move through the rehabilitative process. A collaborative approach to risk assessment, particularly an approach in which risk factors are thoroughly explained and the client contributes to identification of their most relevant treatment needs, provides clients with a sense that they have some control over their assessment and subsequent treatment, in contrast to feeling that assessment and intervention are something done "to them" (Attrill & Liell, 2007; Shingler & Mann, 2006). A structured approach to matching client risk and needs to treatment level can contribute to a sense of "fairness" within risk assessment, which is another area identified as important to offenders (Attrill & Liell, 2007).

While little research has examined offenders' perceptions of risk assessment, Attrill and Leill (2007) interviewed 60 adult sexual offenders regarding their views of risk assessment. A consistent finding during these discussions was offenders' concerns about the level of skill and training of the professionals completing the assessments. The identification of relevant risk factors that are empirically related to recidivism combined with the defined weighting of those risk factors offers an advantage to actuarial measures in this respect. The structured system for weighting the items can make it clear to the client that the assessor's personal biases, level of experience, and skills do not directly influence the assessed level of risk. This is in contrast to SPJ tools that encourage professionals to rely on their experience and skills to examine the risk factors present and determine an overall risk level without a specific structure for combining risk factors (Skeem & Monahan, 2011). The structure associated with actuarial tools, however, has the potential to provide offenders with some sense of consistency, transparency, and evenhandedness to the outcome regardless of the real or perceived qualifications of the assessor.

Critics of actuarial measures note that the specified structure of actuarial tools necessarily limits the "individuality" of risk assessments; this concern was voiced by offenders themselves (Attrill & Liell, 2007). However, as noted earlier in this chapter, the move in recent years toward the integration of dynamic risk assessment with static risk factors provides room for individualization within the overall risk assessment while maintaining the consistency necessary for defensible integrity in implementation. Further, dynamic risk factors allow for more attention to some positive attributes or strengths, which may foster the therapeutic relationship and help in establishing approach rather than avoidance treatment goals (Mann, Webster, Schofield, & Marshall, 2004). As such we would argue that actuarial tools, when implemented well, have the advantage of providing the structure and consistency necessary for strong program integrity and limiting variability in assessor experience and knowledge while still allowing for individuality in the overall assessment.

The Program Level

As described in the first half of this chapter, actuarial risk assessment has evolved as an alternative to UCJ, widely recognized as less accurate, unreliable, and non-replicable. In fact, concerns about the predictive validity of clinical judgment have resulted in the mandated use of actuarial measures within some organizations (e.g., SIR-R used at intake within Correctional Services of Canada; Structured Assessment of Risk and Need, HM Prison Service) and legal jurisdictions. Critics of UCJ note that given its subjective nature, it is difficult to standardize judgments made by a single clinician over time let alone to standardize judgments made by multiple clinicians within one setting. Larger practices and organizations that employ multiple clinicians are often faced with considerable variability in terms of prior training and experience among staff. In more isolated or rural areas, clinicians may be called upon to provide assessments on rare occasions, meaning they bring limited knowledge and expertise to the assessment. The experience and knowledge required to appropriately and effectively use structured professional judgment tools may simply not exist or be realistic in these circumstances.

An advantage of actuarial measures as previously stated is they provide clear direction regarding not only the relevant factors but how to combine those factors into an assessed risk level. Within an intervention program, the detailed manuals that come with many actuarial tools contribute to consistency in application, potentially serve as a guide against which assessments can be audited, provide a training base for new employees, and can minimize program "drift" that may otherwise occur when clinicians are left to make decisions without structured direction. Not only do the manuals associated with actuarial measures provide a framework for appropriate training and skill acquisition for clinicians involved in an evidence-based program, but clinicians report enhanced confidence in their assessments based on actuarial measures (Dr. A. Schweighofer personal communication, 2014). In Neal and Grisso's (2014) survey, the second most common reason cited by psychologists and psychiatrists for using structured tools in risk assessment after ensuring an evidence-based method was "to improve the credibility of my assessment." The third most common reason was "to standardize the assessment" indicating that clinicians themselves perceive value to ensuring that risk assessments have consistent meaning across clinicians, sites, and organizations. Thus there are substantial advantages to the inclusion of actuarial tools in terms of training, consistency, and implementation integrity within evidence-based programs.

The Organizational Level

Leschied, Bernfeld, and Farrington (2001) note that there is political and sometimes philosophical opposition to "what works" in effective correctional interventions. Managerial doubts can undermine the impact and effectiveness of a program.

A good defense to this is to rely on tools with heavy empirical support and demonstrated consistency and replicability; this leaves less room for argument. Actuarial risk assessment meets four goals critical to any organization managing offenders: (1) they identify the level of risk for an individual within a group or population of individuals, (2) they identify contributing salient risk factors that are appropriate targets for intervention (assuming dynamic risk assessment is used), (3) they identify strategies that manage or minimize risk, and (4) they communicate risk information (Mills, Kroner, & Morgan, 2011).

The identification of risk level within a population, along with the contributing risk factors, appropriate treatment targets, and strategies for managing risk, has the potential to directly impact policy in relation to management and intervention of offenders within an organization. A clear management framework and consistent structure for handling offenders within an organization (based on their risk assessment results) should result in time and resource efficiencies. Additionally, a standardized approach facilitates the identification of, planning for, and streamlining of staff training needs. Good quality staff development and training along with subsequent supervision can balance inequality in prior qualifications, knowledge, and skill level among staff members (Mann, Fernandez, & Ware, 2011).

Additionally, the fourth goal of risk communication is critical to the ethical and appropriate management of offenders within an organization. Mills and Kroner (2006) found that risk judgments given using high, moderate, and low categorizations were overestimated, even when the base rate of offending was provided. They note that subjective risk categories lack "solid empirical meaning" and may cause under- or over-estimates of risk, resulting in suboptimal resource allocation to offenders managed within the organization. As noted earlier, actuarial measures typically provide multiple methods to quantify risk, including recidivism estimates, percentiles, and risk ratios along with nominal risk categories. Thus an advantage to actuarial measures is that they provide a common language for risk communication. With appropriate training risk communication will hold the same meaning for everyone within the organization, including decision makers, and directly impact resource allocation.

The Societal Level

Controversy in the use of actuarial risk assessment has focused primarily on its use for decisions related to incarceration (e.g., civil commitment) and release (e.g., parole). There is less controversy over the use of risk assessment as part of treatment planning or about the identification of treatment needs using dynamic risk assessment measures, as is primarily discussed above. There is very little empirical research on the consumption of actuarial risk estimates generally (Scurich, Monahan, & John, 2012). Identified concerns include that decision makers may be misled to think that actuarial tools are more precise than they in fact

are (Campbell, 2007) and consequently overly or inappropriately influence deci-
sions made that impact offenders' lives directly. However, this concern does not
appear to be supported in recent research. For example, offenders referred for full
SVP evaluations tend to have higher risk-measure scores than those who are not
referred; mental health evaluators are more likely to conclude that an offender
meets the criteria for civil commitment when risk scores are high; and attorneys
are more likely to select cases for trial when risk measures are high (Boccaccini
et al., 2009; Levenson, 2004; Murrie, Boccaccini, Rufino, & Caperton, 2012) sug-
gesting that actuarial risk scores play an appropriate and essential role in deter-
mining who are the judges and jurors eventually (see Boccaccini, Turner, Murrie,
Henderson, and Chevalier, 2013).

Once at trial, however, research suggests that mock jurors asked to make deci-
sions in SVP cases are more likely influenced by testimony based on clinical judg-
ment than risk assessment instruments and do not perceive actuarial testimony to be
any more scientific than clinical testimony (Krauss, McCabe, & Lieberman, 2012;
McCabe, Krauss, & Lieberman, 2010). Similarly, Boccaccini et al. (2013) found
that risk-measure scores had little impact on real jurors surveyed after trial in Texas
SVP cases. The authors posited that jurors may perceive that most offenders who are
eligible for SVP commitment (most of whom are identified through actuarial mea-
sures) are "dangerous enough" or that jurors have retributive motives rather than
being concerned with "protecting the public." Regardless of the explanation, it
appears that the use of actuarial measures serves an important purpose at the front
end of this process (i.e., helping to ensure that the most restrictive measures are
applied to the higher risk offenders) while idiosyncratic features may have more
influence during the actual legal proceedings. Neal and Grisso (2014) make the
interesting argument that current forensic training that encourages a too flexible
approach to assessment may be a liability in that it interferes with the ability of
courts to appropriately use risk assessment information as they are "required to
become familiar with a bewilderingly wide range of tools" (p. 1417). The authors
suggest that this could be minimized if clinicians are trained to select tools that are
both appropriate to the referral question and have the best psychometric properties.

As we have noted previously, to be valuable risk assessment results must be
communicated in a clear and appreciable manner to consumers (Heilbrun, Dvoskin,
Hart, & McNeil, 1999). A reliable and valid risk assessment is of no use and in fact
may be "worse than useless" if decision makers misapprehend the results (Heilbrun
et al., 1999, p. 94). Interestingly, one study found that "unpacking" actuarial vio-
lence (i.e., explicitly articulating the extent to which individual risk factors impact
the overall risk) appeared to aid subjects identified as "innumerate" with interpret-
ing the results of actuarial risk assessments and more effectively applying the
group-level risk estimates to the individual case (Scurich et al., 2012). Given the
stakes involved in legal dispositions, we would argue that experts have a particular
ethical obligation when communicating actuarial risk assessment results in high-
stakes circumstances to precede the sharing of results with appropriate education
on the meaning of risk.

Conclusions

While controversy remains regarding the use of actuarial risk assessment, actuarial measures continue to provide the most accurate available information, including for legal decision-making (Heilbrun, 1997). Critics of actuarial tools argue that because actuarial tools do not account for individual differences within their schemes, clinicians are unable to modify level of risk based on mitigating factors, and therefore there is a substantial margin of error inherent in actuarial measures (Hart & Cooke, 2013; Hart, Michie, & Cooke, 2007). Please note, however, that the statistics employed by Hart and colleagues (2007) cannot be used to support their position that group data cannot meaningfully be used to support inferences about individuals (e.g., Hanson & Howard, 2010; G. T. Harris, Rice, & Quinsey, 2008; Mossman & Sellke, 2007; Scurich & John, 2012). Also, it remains to be determined if the posited limitations on individuality produce greater error than clinical overrides based on individual items as applied in SPJ. Further, individuality can be incorporated (at least to some extent) into risk assessment by adding actuarial measures of dynamic risk factors and ensuring that the risk assessment process involves collaboration with the offender. Good risk assessment should use risk estimates obtained by actuarial methods and implemented with integrity, as an "anchor" alongside other measures that include factors that would allow for risk management. Actuarial measures do not replace a clinician's integration and synthesis of information and selection and implementation of a plan of therapeutic action; rather, they can contribute to each aspect of the process. In other words, "scoring an actuarial risk tool is not a risk assessment" (Hanson, 2009, p. 174).

While some of the advantages of actuarial scales described in the present chapter are currently being realized, not all of them are necessarily being maximized by clinicians, programs, or organizations where actuarial risk measures are implemented. When asked, offenders often report a poor understanding of risk assessment, the benefits to them, and little sense of control or impact on the process (Attrill & Liell, 2007). Further, although many newer risk assessment tools include dynamic risk factors, there continues to be a lack of focus on strength or protective factors (Wilson, Desmarais, Nicholls, & Brink, 2010) in the risk assessment process. Wilson et al. note that strengths are not just the opposite of deficits, but capture unique information. This appears to be the next step in risk assessment research.

We also acknowledge that while the importance of consistency and reliability in risk assessment cannot be overemphasized, actuarial measures work best when the offender being assessed possesses characteristics similar to the development sample or validation research of the measure. Regardless of the measure chosen (whether by the clinician or as part of a standardized program or mandated by legislation), it is up to the clinician to ensure that measures used are appropriate to the client being assessed. Actuarial measures are not appropriately applied to every client, and there are circumstances where the current state of the research means that clinical judgment remains the only option. However, in the majority of cases, anchored risk assessment as part of a comprehensive "case conceptualization" should be used to

inform intervention at a more individualized level. In our estimation, the integrated-actuarial approach to risk assessment, when implemented with thought and integrity, holds some valuable clinical advantages while leaving sufficient room for individualization.

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., … Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-Six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382. doi:10.1177/0011000005285875

Aiken, L. R. (1985). *Psychological testing and assessment* (5th ed.). Newton, MA: Allyn and Bacon.

Alexander, J., & Austin, J. (1992). *Handbook for evaluating objective prison classification systems*. San Francisco, CA: National Council on Crime and Delinquency.

Alison, L. J., & Stein, K. L. (2001). Vicious circles: Accounts of stranger sexual assault reflect abusive variants of conventional interactions. *The Journal of Forensic Psychiatry, 12*(3), 515–538. doi:10.1080/09585180127391.

Altman, D. G., Vergouwe, Y., Royston, P., & Moons, K. G. M. (2009). Prognosis and prognostic research: Validating a prognostic model. *BMJ [British Medical Journal], 338*, 1432–1435. doi:10.2307/25671796.

American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist, 68*, 7–19.

Anderson, D., & Hanson, R. K. (2010). Static-99: An actuarial tool to assess risk of sexual and violent recidivism among sexual offenders. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 251–267). New York, NY: Taylor & Francis.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. Cincinnati, OH: Anderson Publishing Co.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52*, 7–27. doi:10.1177/0011128705281756.

Andrews, D. A., & Dowden, C. (2005). Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. *Journal of Legal and Criminological Psychology, 10*, 173–187. doi:10.1348/135532505X36723.

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology, 28*(3), 369–404.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84–94.

Association for the Treatment of Sexual Abusers. (2014). *ATSA practice guidelines for assessment, treatment interventions, and management strategies for male adult sexual abusers*. Beaverton, OR: Professional Issues Committee, ATSA.

Attrill, G., & Liell, G. (2007). Offenders' views on risk assessment. In N. Padfield (Ed.), *Who to release? Parole, fairness and criminal justice* (pp. 191–201). Cullompton, UK: Willan.

Austin, J., Johnson, K. D., & Weitzer, R. (2005). *Alternatives to the secure detention and confinement of juvenile offenders*. Washington, DC: Office of Justice Programs.

Babchishin, K. M. (2013). *Sex offenders do change on risk-relevant propensities: evidence from a longitudinal study of the ACUTE-2007* (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses Global. (MR60297)

Babchishin, K. M., Hanson, R. K., & Blais, J. (2015). A brief scale for predicting violent and general recidivism among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment.* Advance online publication. doi:10.1177/1079063215569544

Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012a). Communicating risk for sex offenders: Risk ratios for Static-2002R. *Sexual Offender Treatment, 7*(2), 1–12.

Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012b). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment, 19*, 442–461. doi:10.1177/1073191112458312.

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

Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse: A Journal of Research and Treatment, 22*, 234–251. doi:10.1177/1079063210369014.

Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology, 79*, 182–192. doi:10.1037/a0022900.

Bennell, C., Alison, L. J., Stein, K. L., Alison, E. K., & Canter, D. V. (2001). Sexual offenses against children as the abusive exploitation of conventional adult-child relationships. *Journal of Social and Personal Relationships, 18*(2), 155–171. doi:http://dx.doi.org/10.1177/0265407501182001.

Bernfeld, G. A., Blase, K. A., & Fixsen, D. L. (1990). Towards a unified perspective on human service delivery systems: Application of the Teaching-Family Model. In R. J. McMahon & R. D. Peters (Eds.), *Behavior disorders of adolescents: Research, intervention and policy in clinical and school settings* (pp. 191–205). New York, NY: Plenum.

Blais, J., & Forth, A. E. (2014). Prosecution-retained versus court-appointed experts: Comparing and contrasting risk assessment reports in preventative detention hearings. *Law and Human Behavior, 38*, 531–543. doi:10.1037/lhb0000082.

Boccaccini, M. T., Murrie, D. C., Caperton, J. D., & Hawes, S. W. (2009). Field validity of the Static-99 and MnSOST-R among sex offenders evaluated for civil commitment as sexually violent predators. *Psychology, Public Policy, and Law, 15*, 278–314. doi:10.1037/a0017232.

Boccaccini, M. T., Turner, D. B., Murrie, D. C., Henderson, C. E., & Chevalier, C. (2013). Do scores from risk measures matter to jurors? *Psychology, Public Policy, and Law, 19*, 259–269. doi:10.1037/a0031354.

Boer, D. P., Wilson, R. J., Gauthier, C. M., & Hart, S. D. (1997). Assessing risk of sexual violence: Guidelines for clinical practice. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 326–342). New York, NY: Guilford Press.

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18–32). Thousand Oaks, CA: Sage.

Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice & Behavior, 29*(4), 355–379.

Bonta, J., Law, M., & Hanson, R. K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*, 123–142. doi:10.1037/0033-2909.123.2.123.

Bonta, J., & Motiuk, L. L. (1990). Classification to halfway houses: A quasi-experimental evaluation. *Criminology, 23*(3), 497–506.

Brouillette-Alarie, S., Babchishin, K. M., Hanson, R. K., & Helmus, L. M. (2015). Latent constructs of static risk scales for the prediction of sexual recidivism: A 3-factor solution. *Assessment*. Advance online publication. doi:10.1177/1073191114568114

Campbell, T. W. (2007). *Assessing sex offenders: Problems and pitfalls* (2nd ed.). Springfield, IL: Charles Thomas.

Canter, D. V. (1994). *Criminal shadows*. London: Harper Collins.

Canter, D. V. (2004). Offender profiling and investigative psychology. *Journal of Investigative Psychology and Offender Profiling, 1*(1), 1–15. doi:10.1002/jip.7.

Canter, D. V., Bennell, C., Alison, L. J., & Reddy, S. (2003). Differentiating sex offences: A behaviorally based thematic classification of stranger rapes. *Behavioral Sciences & the Law, 21*(2), 157–174. doi:http://dx.doi.org/10.1002/bsl.526.

Canter, D. V., & Heritage, R. (1990). A multivariate model of sexual offence behaviour: Developments in 'offender profiling'. *The Journal of Forensic Psychiatry, 1*(2), 185–212. doi:10.1080/09585189008408469.

Canter, D. V., Hughes, D., & Kirby, S. (1998). Paedophilia: Pathology, criminality, or both? The development of a multivariate model of offence behaviour in child sexual abuse. *The Journal of Forensic Psychiatry, 9*(3), 532–555. doi:10.1080/09585189808405372.

Chevalier, C., Boccaccini, M. T., Murrie, D. C., & Varela, J. G. (2014). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior*. Advance online publication. doi:10.1037/lhb0000114

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104. doi:10.1037//0021-9010.78.1.98.

Crawford, J. R., & Garthwaite, P. H. (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks. *The Clinical Neuropsychologist, 23*, 193–204. doi:10.1080/13854040801968450.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. doi:10.1007/BF02310555.

Dahle, K.-P., Biedermann, J., Lehmann, R. J. B., & Gallasch-Nemitz, F. (2014). The development of the crime scene behavior risk measure for sexual offense recidivism. *Law and Human Behavior, 38*, 569–579. doi:10.1037/lhb0000088.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1674. doi:10.1126/science.2648573.

Doren, D. M. (2004). Stability of the interpretative risk percentages for the RRASOR and Static-99. *Sexual Abuse: A Journal of Research and Treatment, 16*, 25–36. doi:10.1177/107906320401600102.

Dowden, D., & Andrews, D. A. (2004). The importance of staff practice in delivering effective correctional treatment: A meta-analytic review of core correctional practice. *International Journal of Offender Therapy and Comparative Criminology, 48*, 203–214. doi:10.1177/0306624X03257765.

Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment, 24*, 5–28. doi:10.1177/1079063211403164.

Elmore, J. G., & Gigerenzer, G. (2005). Benign breast disease: The risk of communicating risk. *New England Journal of Medicine, 353*, 297–299. doi:10.1056/NEJMe058111.

Farrington, D. P., & Ttofi, M. M. (2011). Protective and promotive factors in the development of offending. In T. Bliesener, A. Beelmann, & M. Stemmler (Eds.), *Antisocial behavior and crime: Contributions of developmental and evaluation research to prevention and intervention* (pp. 71–88). Cambridge, MA: Hogrefe Publishing.

Fernandez, Y. (2008, October). *An examination of the inter-rater reliability of the Static-99 and STABLE-2007.* Poster presented at the 27th Annual Conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014). *STABLE-2007 coding manual – revised 2014*. Unpublished report. Ottawa, ON: Public Safety Canada.

Flores, A. W., Russell, A. L., Latessa, E. J., & Travis, L. F. (2005). Evidence of professionalism or quackery: Measuring practitioner awareness of risk/need factors and effective treatment strategies. *Federal Probation, 69*, 9–14.

Gail, M. H., & Pfeiffer, R. M. (2005). On criteria for evaluating models of absolute risk. *Biostatistics, 6*, 227–239. doi:10.1093/biostatistics/kxi005.

Goodwill, A. M., Alison, L. J., & Beech, A. R. (2009). What works in offender profiling? A comparison of typological, thematic, and multivariate models. *Behavioral Sciences & the Law, 27*, 507–529. doi:10.1002/bsl.867.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30. doi:10.1037//1040-3590.12.1.19.

Häkkänen, H., Lindlöf, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in a sample of Finnish stranger rapes. *Journal of Investigative Psychology and Offender Profiling, 1*(1), 17–32. doi:10.1002/jip.1.

Häkkänen, H., Puolakka, P., & Santtila, P. (2004). Crime scene actions and offender characteristics in arsons. *Legal and Criminological Psychology, 9*(2), 197–214. doi:10.1348/1355325041719392.

Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence, 20*, 212–217. doi:10.1177/0886260504267740.

Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology, 50*, 172–182. doi:10.1037/a0015726.

Hanson, R. K., Babchishin, K. M., Helmus, L., & Thornton, D. (2013). Quantifying the relative risk of sex offenders: Risk ratios for Static-99R. *Sexual Abuse: A Journal of Research and Treatment, 25*, 482–515. doi:10.1177/107906321246906.

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348–362. doi:10.1037/0022-006X.66.2.348.

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*, 6–35. doi:10.1177/0093854800 027001002.

Hanson, R. K., & Harris, A. J. R. (2013). Criminogenic needs of sexual offenders on community supervision. In L. A. Craig, L. Dixon, & T. A. Gannon (Eds.), *What works in offender rehabilitation: An evidenced-based approach to assessment and treatment* (pp. 421–435). Chichester, UK: Wiley-Blackwell.

Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. Ottawa, ON: Public Safety Canada.

Hanson, R. K., Helmus, L.-M., & Harris, A. J. R. (2015). Assessing the Risk and Needs of Supervised Sexual Offenders: A Prospective Study Using STABLE-2007, Static-99R, and Static-2002R. *Criminal Justice and Behavior, 42*(12), 1205–1224. doi:10.1177/0093854815602094.

Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism among sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior, 34*, 198–211. doi:10.1007/s10979-009-9180-1.

Hanson, R. K., & Howard, P. D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior, 34*, 275–281. doi:10.1007/s10979-010-9227-3.

Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2012). Developing non-arbitrary metrics for risk communication: Percentile ranks for the Static-99/R and Static-2002/R sexual offender risk tools. *International Journal of Forensic Mental Health, 11*, 9–23. doi:10.1080/14999013. 2012.667511.

Hanson, R. K., Lunetta, A., Phenix, A., Neeley, J., & Epperson, D. (2014). The field validity of Static-99/R sex offender risk assessment tool in California. *Journal of Threat Assessment and Management, 1*, 102–117. doi:10.1037/tam0000014.

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154–1163. doi:10.1037/0022-006X.73.6.1154.

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*, 1–21. doi:10.1037/a0014421.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119–136. doi:10.1023/A:1005482921333.

Hanson, R. K., Thornton, D., Helmus, L., & Babchishin, K. M. (2015). What sexual recidivism rates are associated with Static-99R and Static-2002R scores? *Sexual Abuse: A Journal of Research and Treatment.* Advance online publication. doi:10.1177/1079063215574710

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised technical manual* (2nd ed.). Toronto, ON, Canada: Multi-Health Systems.

Harris, A. J. R., & Hanson, R. K. (2010). Clinical, actuarial and dynamic risk assessment of sexual offenders: Why do things keep changing? *Journal of Sexual Aggression, 16*, 296–310. doi:10.1080/13552600.2010.494772.

Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99: Coding rules revised 2003*. Ottawa, ON: Solicitor General Canada.

Harris, G. T., & Rice, M. E. (2015). Progress in violence risk appraisal and communication: A commentary on hypotheses and evidence. *Behavioral Sciences and the Law, 33*, 128–145. doi:10.1002/bsl.2157.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413–425.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (2008). Shall evidence-based risk appraisal be abandoned? *British Journal of Psychiatry, 192,* 154. (expanded version at http://bjp.rcpsych.org/cgi/eletters/190/49/s60#5674).

Hart, S. D., & Cooke, D. J. (2013). Another look at the (im-) precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences & the Law, 31*, 81–102. doi:10.1002/bsl.2049.

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry, 190*, s60–s65.

Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior, 21*, 347–359. doi:10.1023/A:1024851017947.

Heilbrun, K., Dvoskin, J., Hart, S., & McNeil, D. (1999). Violence risk communication: Implications for research, policy, and practice. *Health, Risk & Society, 1*, 91–106.

Helmus, L., Babchishin, K. M., & Hanson, R. K. (2013). The predictive accuracy of the Risk Matrix 2000: A meta-analysis. *Sexual Offender Treatment, 8*(2), 1–24.

Helmus, L., & Hanson, R. K. (2013). *STABLE-2007: Updated recidivism rates* (includes combinations with Static-99R, Static-2002R, and Risk Matrix 2000). Unpublished report. Ottawa, ON: Public Safety Canada.

Helmus, L., Hanson, R. K., Babchishin, K. M., & Thornton, D. (2014). Sex offender risk assessment with the Risk Matrix 2000: Validation and guidelines for combining with the STABLE-2007. *Journal of Sexual Aggression*. Advance online publication. doi:10.1080/13552600.2013.870241

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior, 39*, 1148–1171. doi:10.1177/0093854812443648.

Helmus, L., & Thornton, D. (2014). The MATS-1 risk assessment scale: Summary of methodological concerns and an empirical validation. *Sexual Abuse: A Journal of Research and Treatment.* Advance online publication. doi:10.1177/1079063214529801

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment, 24*, 64–101. doi:10.1177/1079063211409951.

Hilton, N. Z. (2014). Actuarial assessment in serial intimate partner violence: Comment on Cook, Murray, Amat and Hart. *Journal of Threat Assessment and Management, 1*, 87–92. doi:10.1037/tam0000013.

Hilton, N. Z., Carter, A. M., Harris, G. T., & Sharpe, A. J. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *Journal of Interpersonal Violence, 23*, 171–188.

Hilton, N. Z., Scurich, N., & Helmus, L. M. (2015). Communicating the risk of violent and offending behavior: Review and introduction to special issue. *Behavioral Sciences and the Law, 33*, 1–18. doi:10.1002/bsl.2160.

Hunsley, J. D., & Mash, E. J. (2010). The role of assessment in evidence-based practice. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (2nd ed.). New York, NY: Guilford Press.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15*, 446–455. doi:10.1037/1040-3590.15.4.446.

Interstate Commission for Adult Offender Supervision. (2007). *Sex offender assessment information survey.* (ICAOS Documents No. 4-2007).

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment, 19*, 425–448. doi:10.1177/107906320701900407.

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Macmillan.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*, 515–526. doi:10.1037/a0016755.

Krauss, D. A., McCabe, J., & Lieberman, J. (2012). Dangerously misunderstood: Representative jurors' reactions to expert testimony on future dangerousness in a sexual violent predator trial. *Psychology, Public Policy, and Law, 18*, 18–49.

Kropp, P. R., & Gibas, A. (2010). The Spousal Assault Risk Assessment Guide (SARA). In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 227–250). New York, NY: Routledge.

Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) Guide: Reliability and validity in adult male offenders. *Law and Human Behavior, 24*, 101–118. doi:10.1023/A:1005430904495.

Lehmann, R. J. B., Goodwill, A. M., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K.-P. (2013). Applying crime scene analysis to the prediction of sexual recidivism in stranger rapes. *Law and Human Behavior, 37*, 241–254. doi:10.1037/lhb0000015.

Lehmann, R. J. B., Goodwill, A. M., Hanson, R. K., & Dahle, K.-P. (2014). Crime scene behaviors indicate risk-relevant propensities of child molesters. *Criminal Justice and Behavior, 41*(8), 1008–1028. doi:10.1177/0093854814521807.

Lehmann, R. J. B., Goodwill, A. M., Hanson, R. K., & Dahle, K.-P. (2015). *Acquaintance rape: Applying crime scene analysis to the prediction of sexual recidivism*. Sexual Abuse: A Journal of Research and Treatment. Advance online publication. doi:10.1177/1079063215569542.

Lehmann, R. J. B., Hanson, R. K., Babchishin, K. M., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K.-P. (2013). Interpreting multiple risk scales for sex offenders: Evidence for averaging. *Psychological Assessment, 25*, 1019–1024. doi:10.1037/a0033098.

Lehmann, R. J. B., Thornton, D., Helmus, L. M., & Hanson, R. K. (2015). *Developing non-arbitrary metrics for risk communication: Norms for the Risk Matrix 2000.* Unpublished manuscript.

Leschied, A. W., Bernfeld, G. A., & Farrington, D. P. (2001). Implementation issues. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice: Implementing and evaluating effective programs* (pp. 3–24). Chichester, UK: Wiley.

Levenson, J. S. (2004). Sexual predator civil commitment: A comparison of selected and released offenders. *International Journal of Offender Therapy and Comparative Criminology, 48*, 638–648. doi:10.1177/0306624X04265089.

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine, 43*(suppl 1), S8–S23. doi:10.1016/j.amepre.2012.04.029.

Lowenkamp, C., Pealer, J., Smith, P., & Latessa, E. J. (2006). Adhering to the risk and need principles: Does it matter for supervision-based programs? *Federal Probation, 70*, 3–8.

Mann, R., Fernandez, Y., & Ware, J. (2011, November). *Managing sex offender treatment programmes: A professional development workshop for managers and supervisors.* Full day pre-conference workshop presented at the Association for the Treatment of Sexual Abusers, 30th Annual Research and Treatment Conference, Toronto, ON, Canada.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment, 22*, 191–217. doi:10.1177/1079063210366039.

Mann, R. E., Webster, S. D., Schofield, C., & Marshall, W. L. (2004). Approach versus avoidance goals in relapse prevention with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 16*, 65–75. doi:10.1023/B:SEBU.0000006285.73534.57.

McCabe, J., Krauss, D. A., & Lieberman, J. (2010). Reality check: A comparison of college students and community samples of mock jurors in a simulated sexual violent predator civil commitment. *Behavioral Sciences & the Law, 28*, 730–750. doi:10.1002/bsl.902.

McGrath, R. J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, E. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey*. Brandon, VT: Safer Society Press.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Mills, J. F., Jones, M. N., & Kroner, D. G. (2005). An examination of the generalizability of the LSI-R and VRAG probability bins. *Criminal Justice and Behavior, 32*, 565–585. doi:10.1177/0093854805278417.

Mills, J. F., & Kroner, D. G. (2006). The effect of base-rate information on the perception of risk for reoffense. *American journal of forensic psychology, 24*(3), 45–56.

Mills, J. F., Kroner, D. G., & Morgan, R. D. (2011). *Clinician's guide to violence risk assessment*. New York, NY: Guildford Press.

Mokros, A. (2007). *Die Struktur der Zusammenhänge von Tatbegehungsmerkmalen und Persönlichkeitseigenschaßen bei Sexualstraftätem* [The structure of the relationship between crime scene actions and personality characteristics in sex offenders]. Frankfurt: Verlag für Polizeiwissenschaft.

Monahan, J., & Silver, E. (2003). Judicial decision thresholds for violence risk management. *International Journal of Forensic Mental Health, 2*(1), 1–6.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783–792. doi:10.1037/0022-006X.62.4.783.

Mossman, D. & Sellke, T. (2007). Avoiding errors about "margin of error." *British Journal of Psychiatry;* Electronic letter in response to Hart, Michie, & Cooke, 2007.

Murrie, D. C., Boccaccini, M. T., Rufino, K., & Caperton, J. (2012). Field validity of the Psychopathy Checklist-Revised in sex offender risk assessment. *Psychological Assessment, 24*, 524–529. doi:10.1037/a0026015.

National Policing Improvement Agency. (2010). *Guidance on protecting the public: Managing sexual offenders and violent offenders*. Retrieved from http://www.acpo.police.uk/documents/crime/2010/20110301%20CBA%20ACPO%20%282010%29%20Guidance%20on%20Protecting%20the%20Public%20v2%20main%20version.pdf.

Neal, T. M. S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior, 41*, 1406–1421. doi:10.1177/0093854814548449.

Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. P. (2014). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale – Sexual Offender version. *Sexual Abuse: A Journal of Research and Treatment, 26*, 472–499. doi:10.1177/1079063213502679.

Olver, M. E., Beggs Christofferson, S. M., & Wong, S. C. P. (2015). Evaluation and applications of the Clinically Significant Change method with the Violence Risk Scale - Sexual Offender version: Implications for risk-change communication. *Behavioral Sciences and the Law, 33*, 92–110. doi:10.1002/bsl.2159.

Olver, M. E., Wong, S. C. P., Nicholaichuk, T. P., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318–329. doi:10.1037/1040-3590.19.3.318.

Pedneault, A. (2014). *Linking crime scene behaviors and propensities in child molesters: A replication.* Paper presented at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers (ATSA), San Diego, CA.

Phenix, A., & Epperson, D. (2015). Overview of the development, reliability, validity, scoring, and uses of the Static-99, Static-99R, Static-2002, and Static-2002R. In A. Phenix & H. M. Hoberman (Eds.), *Sexual offending: Predisposing conditions, assessments, and management.* Springer.

Phenix, A., Helmus, L., & Hanson, R. K. (2015). *Static-99R & Static-2002R evaluator's workbook*. Retrieved from http://www.static99.org/pdfdocs/Static-99RandStatic-2002R_Evaluators Workbook2012-07-26.pdf.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 1102–1114. doi:10.1037/ 0003-066X.47.9.1102.

Proulx, J., Tardiff, M., Lamoureeux, B., & Lussier, P. (2000). How does recidivism risk assessment predict survival? In D. R. Laws, S. M. Hudson, & T. Ward (Eds.), *Remaking relapse prevention with sex offenders: A sourcebook*. Thousand Oaks, CA: Sage Publications.

Quesada, S. P., Calkins, C., & Jeglic, E. L. (2013). An examination of the interrater reliability between practitioners and researchers on the Static-99. *International Journal of Offender Therapy and Comparative Criminology, 58*, 1364–1375. doi:10.1177/0306624X13495504.

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

Rockhill, B., Byrne, C., Rosner, B., Louie, M. M., & Colditz, G. (2003). Breast cancer risk prediction with a log-incidence model: Evaluation of accuracy. *Journal of Clinical Epidemiology, 56*, 856–861. doi:10.1016/S0895-4356(03)00124-0.

Santtila, P., Häkkänen, H., Canter, D., & Elfgren, T. (2003). Classifying homicide offenders and predicting their characteristics from crime scene behavior. *Scandinavian Journal of Psychology, 44*(2), 107–118. doi:10.1111/1467-9450.00328.

Scurich, N., & John, R. S. (2012). A Bayesian approach to the group versus individual prediction controversy in actuarial risk assessment. *Law and Human Behavior, 36*, 237–246. doi:10.1037/ h0093973.

Scurich, N., Monahan, J., & John, R. S. (2012). Innumeracy and unpacking: Bridging the nomothetic/idiographic divide in violence risk assessment. *Law and Human Behavior, 36*, 548–554. doi:10.1037/h0093994.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252–262. doi:10.1016/0749-5978(92)90064-E.

Shingler, J., & Mann, R. E. (2006). Collaboration in clinical work with sexual offenders: Treatment and assessment. In W. L. Marshall, Y. Fernandez, L. Marshall, & G. Serran (Eds.), *Sexual offender treatment: Controversial issues*. London: Wiley & Sons, Ltd.

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., … Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health, 13*, 193–206. doi:10.1080/14999013.2014.922141

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*, 38–42. doi:10.1177/0963721410397271.

Smith, P., & Schweitzer, M. (2012). The therapeutic prison. *Journal of Contemporary Criminal Justice, 28*, 7–22. doi:10.1177/1043986211432201.

Snowden, R. J., Gray, N. S., Taylor, J., & MacCulloch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine, 37*, 1539–1549. doi:10.1017/S0033291707000876.

Social Work Inspection Agency, HM Inspectorate of Constabulary for Scotland, and HM Inspectorate of Prisons. (2009). *Multi-agency inspection: Assessing and managing offenders who present a high risk of serious harm 2009*. Retrieved from http://www.scotland.gov.uk/ Resource/Doc/275852/0082871.pdf

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1–26. doi:10.1111/1529-1006.001.

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. *Annals of the New York Academy of Sciences, 989*(1), 225–235. doi:http://dx.doi.org/10.1111/j.1749-6632.2003.tb07308.x

Varela, J. G., Boccaccini, M. T., Cuervo, V. A., Murrie, D. C., & Clark, J. W. (2014). Same score, different message: Perceptions of offender risk depend on Static-99R risk communication format. *Law and Human Behavior, 38*, 418–427. doi:10.1037/lhb0000073.

Viallon, V., Ragusa, S., Clavel-Chapelon, F., & Bénichou, J. (2009). How to evaluate the calibration of a disease risk prediction tool. *Statistics in Medicine, 28*, 901–916. doi:10.1002/sim.3517.

Ward, T., Polaschek, D. L. L., & Beech, A. R. (2005). *Theories of sexual offending*. Chichester, England: Wiley.

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. 251–277). New York, NY: Guilford Press.

Webster, S. D., Mann, R. E., Carter, A. J., Long, J., Milner, R. J., O'Brien, M. D., … & Ray, N. (2006). Inter-rater reliability of dynamic risk assessment with sexual offenders. *Psychology, Crime, & Law, 12,* 439–452.

Welsh, J. L., Schmidt, F., McKinnon, L., Chattha, H. K., & Meyers, J. R. (2008). A comparative study of adolescent risk assessment instruments predictive and incremental validity. *Assessment, 15*, 104–115.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., & Brink, J. (2010). The role of client strengths in assessments of short-term violence risk. *International Journal of Forensic Mental Health Services, 9*, 282–293.

Wilson, M. A., & Leith, S. (2001). Acquaintances, lovers, and friends: Rape within relationships. *Journal of Applied Social Psychology, 31*(8), 1709–1726. doi:10.1111/j.1559-1816.2001.tb02747.x.

Wong, S. C. P., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). *The Violence Risk Scale: Sexual Offender version (VRS: SO)*. Saskatoon, Saskatchewan, Canada: Regional Psychiatric Centre and University of Saskatchewan.

Wormith, J. S., Althouse, M. S., Reitzel, L. R., Fagan, T. J., & Morgan, R. D. (2007). The rehabilitation and reintegration of offenders: The current landscape and some future directions for correctional psychology. *Criminal Justice and Behaviour, 34*, 879–892. doi:10.1177/0093854807301552.

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior, 39*, 1511–1538. doi:10.1177/0093854812455741.