Chapter 2
Assessment Instruments

We now have a vast amount of data regarding risk assessments from the point of view of demographic, psychological, sociological, biological, and contextual bases. We also have a variety of different techniques to utilize all the way from the straight (unaided) clinical method, the purely actuarial or adjusted actuarial model, and the structured professional judgment approach. We have judgment calls to make in fields as diverse as violence (in general), domestic violence, and sexual violence. We have groups of professionals writing articles proclaiming their instruments to be superior to all others and, further, that all the others do not meet a "standard of care in the assessment of violent behavior." What are we to do with this dizzying array of approaches, and the claims of each to be superior over the others? Or the fact that manipulation of statistics can actually be used to "prove" the superiority of one method over another, meaning there may not be a truly superior instrument to choose?

With the exception of unaided clinical judgment, which most people do not utilize in modern practice, the various approaches to risk assessment are essentially equivalent in their ability to assess violent behavior. After an extensive and exhaustive review of the different risk assessment approaches, Heilbrun (2009) concluded that carefully used, structured professional judgment is just as accurate as actuarial assessment. It is fascinating that people who favor the actuarial approach frequently contrast their findings with those of unaided clinical judgment, not even acknowledging that there is an alternative technique out there called structured professional judgment. This amounts to beating a dead horse, since no one is claiming that unaided clinical judgment still has utility these days.

Another conceptual issue that should be mentioned has to do with placement of the individual into a risk category, or group. Just because a person is assigned to a certain category based on the results of an evaluation does not necessarily mean s/he will fit the entire pattern, as there may be specific idiographic factors that differentiate this individual. This is, essentially, what the adjusted actuarial approach is about; however, those advocating for the purely actuarial model contend that any type of adjustment is merely a throwback to clinical judgment. Those utilizing adjusted actuarial approaches, however, argue that they use the actuarial model

as an anchoring point and then refine that anchoring point with specific idiographic data. Basically, when we obtain a predictive value for someone in a group, depending on certain other variables, the actual predictive statement may be even higher (or lower). If, for instance, an individual scores in the moderate range on a risk assessment measure but has a mental illness, presents with features of psychopathy and paranoia, has a history of self-medicating with illicit substances, refuses to take his prescribed medication, is unemployed, and has a history of traumatic brain injury, the risk of future violent behavior would more likely be quite high than moderate. Consequently, the importance of modifications to the overall likelihood must be considered.

A final issue worth discussing is that there may not be sufficient data available to make an assessment in any given case. For example, we may not have a complete or factual mental health history, an accurate assessment of the presence of a personality disorder, incomplete records relating to previous violence (or the individual not telling the truth about it), and no records relating to the person's employment status. In the event there is such a scarcity of information, we need to be able to say that there is insufficient data to make a complete assessment.

Let us now look at several risk assessment instruments, both actuarial and those described as structured professional judgment. We will start with the Psychopathy Checklist-Revised (PCL-R) as many of the other tools utilize a PCL-R score as one of their most robust variables.

Psychopathy Checklist-Revised (PCL-R)

The PCL-R can be thought of as an update and combination of the 16 personality traits outlined by Cleckley (1976), who spoke of the psychopath as presenting "a mask of sanity." Robert Hare updated this concept based on his own clinical experience in correctional settings and sought to build on the existing knowledge by developing a system by which to assess for the construct. The resultant Psychopathy Checklist (Hare 1980) has been revised twice since its induction (hence, PCL-R). The first revision of the Psychopathy Checklist (Hare 1991) was limited in its applicability to diverse offender groups. The second edition, described by the author as a "conservative" revision (Hare 2003, p. 1), expanded the standardization data to include a variety of other offenders (e.g., African-American offenders, female offenders, sex offenders, substance abusers, and younger offenders; Hare 2003, p. 2). It is interesting that the PCL-R was not designed with a specific focus on violence risk assessment; yet, its total score is identified as a variable for consideration on many of the available risk assessment instruments.

The PCL-R is composed of 20 items, each of which gauges for an essential behavior pattern or trait inherent in the prototypical psychopath (Rogers and Graves-Oliver 2003). These 20 items are divided into two groups, regarded as Factors, and then each Factor is further subdivided into two Facets. The items in

Factor One assess for the interpersonal (Facet One) and affective (Facet Two) deficits inherent in the construct of psychopathy. Not surprisingly, these variables have been found to be positively correlated with measures of narcissism and negatively correlated with measures of state/trait anxiety (Harpur et al. 1989). The Factor One items include:

(A) Facet One—Interpersonal:

  (a) Glibness, superficial charm;
  (b) Grandiose sense of self-worth;
  (c) Pathological lying; and,
  (d) Conning and manipulative interpersonal style.

(B) Facet Two—Affective:

  (a) Lack of remorse or guilt;
  (b) Shallow affect;
  (c) Callousness and lack of empathy; and,
  (d) Failure to accept responsibility for one's actions.

The items in Factor Two, on the other hand, assess for the lifestyle (Facet Three) and the antisocial (Facet Four) behavior patterns characteristic of the prototypical psychopath. These variables have been found to be positively correlated with measures of sensation-seeking and impulsivity (Harpur et al. 1989). The specific Factor Two items include:

(A) Facet Three—Lifestyle:

  (a) Need for stimulation and proneness to boredom;
  (b) Parasitic lifestyle;
  (c) Lack of realistic long-term goals;
  (d) Impulsivity; and,
  (e) Irresponsibility.

(B) Facet Four—Antisocial:

  (a) Poor behavioral controls;
  (b) Early behavioral problems;
  (c) Juvenile delinquency;
  (d) Revocation of conditional release; and,
  (e) Criminal versatility.

The Factors and Facets that comprise the PCL-R were determined by means of a factor analysis performed to identify the most meaningful combination of the variables possible. There are two additional components not included in the Factors that have been identified as part of the construct of psychopathy, namely Promiscuous Sexual Behavior and Many Short-Term Marital Relationships (Hare 2003).

   Indeed, many of the items listed in Factor One and Factor Two correspond fairly closely to the diagnostic criteria for Antisocial Personality Disorder found

in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision* (*DSM-IV-TR*) (American Psychiatric Association 2000). Psychopathy is not an official diagnosis in the DSM-IV-TR, though, nor does it appear in the DSM-5 (American Psychiatric Association 2013). Therefore, even if someone meets the cutoff score for psychopathy on the PCL-R, they would still be diagnosed as Antisocial Personality Disorder according to the DSM-IV-TR. Some authors (e.g., Hare et al. 1991; Hare 1996) have argued that a diagnosis of psychopathy and a diagnosis of Antisocial Personality Disorder are, in fact, quite distinct. Approximately 60 % of prison inmates meet the DSM-IV-TR criteria for ASPD, but only 20–25 % of that 60 % (approximately 10–15 % of the prison population) endorsed enough of the risk factors to be classified as a psychopath (Hare 2003). This may explain why the Factor One items are considered the core traits of the prototypical psychopath, as most do not overlap with the DSM-IV-TR diagnosis of Antisocial Personality Disorder.

Scoring the PCL-R is both complex and straightforward. A three-point coding system is used to score each item, with a 0 assigned if the risk factor is not present, a 1 if there is some indication that the risk factor is present but not to the degree needed for a full rating of 2, and a score of 2 given if the risk factor is *definitely* present. Each rating is based on the typical functioning throughout the course of the respondent's life (Hare 2003). In other words, the trait must be a stable feature. The highest possible score for the measure is 40, with 30 assigned as the cutoff score by Hare (2003) for a description of psychopathy. In the event that there is limited information available, up to five items may be omitted without invalidating the total score.

The final scores recorded for the PCL-R include a total score for Factor One, a total score for Factor Two, and an overall total score. One should never base these ratings just on self-report and a substantial amount of collateral information is required to do the ratings (Hare 2003, p. 19). Very often, a semi-structured interview is used to give the evaluator an opportunity to observe whether the offender demonstrates any of the factor items during the interaction, such as glibness and superficial charm, grandiosity, and pathological lying. Therefore, both the semi-structured interview and use of collateral data is recommended for completeness.

The PCL-R total score can be viewed as indicative of the degree to which an individual resembles the prototypical psychopath. The general interpretation of the total score on the PCL-R is quite straightforward: A score of 30 or greater indicates the offender can be classified as a psychopath. This categorical option, however, is discouraged by Hare (2003, p. 31) as it is probably not psychometrically defensible. Indeed, Marcus, John, and Edens (2004) have noted that it may be better to think of psychopathy as a continuum rather than a 'yes' or 'no' category. Therefore, a suggested method is to use ranges (Olmi 2005). It is also useful to consider the range of scores on the four Facets to help clarify what makes a particular psychopath unique from others.

While the categorical option may not be psychometrically defensible, it nevertheless provides insight concerning behaviors to expect in individuals who do, in fact, reach the cutoff mark. For example, a total score of 30 or greater may

indicate that the offender is likely to violate the terms of a Conditional Release (Hart et al. 1988). Moreover, a score in the elevated range is strongly associated with poor responsiveness to treatment, indicating that certain interventions are contraindicated. Within the area of interpersonal violence, the confirmation of a psychopathy diagnosis should, ideally, be considered when deciding which interventions are appropriate for a particular individual, as well as which would be contraindicated (Boyle et al. 2008).

There are two alternative versions: the Psychopathy Checklist: Screening Version (PCL:SV) (Hart et al. 1995) and the Psychopathy Checklist: Youth Version (PCL:YV) (Forth et al. 2003). The PCL:SV is a 12 item scale based on a subset of PCL-R items that was created during the MacArthur Risk Assessment studies and is often substituted for the PCL-R to save time. This measure includes more Factor One items than it does Factor Two items. The PCL:YV is helpful for the rapidly growing research concerning the juvenile psychopaths and is distinguished from the PCL-R by its rephrasing of select items to make the respective variables more applicable to juveniles. For instance, the PCL-R item Poor Behavioral Controls appears on the youth version as Poor Anger Control and Revocation of Conditional Release appears as Serious Violation of Conditional Release. The examiner needs to be careful when using this instrument on adolescents, however, because many of the items are common in juvenile behavior (e.g., a need for stimulation, impulsivity, failure to accept responsibility) and may fade over time. Theoretically, this is due to the development of areas of the brain responsible for executive functioning that occurs in the later teen years.

It is important to emphasize that only those appropriately trained should administer this instrument (Hare 1998) as it requires a high degree of sensitivity to ensure that the decision-making style is in line with that used during the tool's standardization.

Violence Risk Appraisal Guide (VRAG)

The VRAG is a series of weighted items meant to determine future risk of violent recidivism in adult males (Quinsey et al. 2006). The early stages of the VRAG's development began with a review of the literature to determine measurable risk factors correlated with violence. After, regression analyses were performed using data from samples of Canadian forensic psychiatric patients to determine which of the identified risk factors were the most strongly predictive of violence (Harris et al. 1993). It should be noted that this was a postdictive, rather than a predictive, study comparing factors that characterized one group as opposed to those which characterized another group (violent versus nonviolent). A book conceptualizing the approach was released in 1998; the second edition was released in 2006.

Overall, Quinsey et al. (1998, 2006) identified 12 variables for consideration during the assessment, namely: (1) Lived with biological parents until the age of 16 (except for death of a parent); (2) Elementary school maladjustment;

(3) History of alcohol problems (each of the following gets one point: Parental alcohol abuse, adult alcohol abuse, alcohol involved in index offense, teenage alcohol problem, alcohol involved in a prior offense); (4) Marital status (married receives a negative weight and unmarried receives a positive weight—that is, more likely to act violently); (5) Criminal history score for nonviolent offenses (22 crimes are listed according to a rating system with the most serious being robbery and arson, going down to less serious crimes such as making threats, breaking and entering, forgery of a check, and disturbing the peace); (6) Failure on prior Conditional Release; (7) Age at index offense; (8) Victim injury (death is scored −2, which is the lowest likelihood because this group would be unlikely to repeat the fatal criminal behavior); (9) Female victim (index offense); (10) Personality disorder (positively weighted); (11) Schizophrenia (negatively weighted); and (12) Psychopathy score. While the authors have recommended that the PCL-R be used for this item, they have also acknowledged that its administration may not always be possible or warranted. To compensate for this potential shortcoming, the authors explored psychopathy as a taxon (Harris et al. 1994) and later incorporated the CATS (Quinsey et al. 1998), which is a brief checklist of eight items designed to measure for traits of psychopathy endorsed by the respondent as a juvenile (e.g., suspensions/expulsions from school, arrested under the age of 16).

Total VRAG scores range from −24 to +38 and correspond to 7- and 10-year probability estimates for violent recidivism (Quinsey et al. 2006). The total score is used in conjunction with the VRAG table to interpret the results. The VRAG table is provided in the instrument's protocol and is broken down into two columns: VRAG score and Category of Risk, which denotes the risk level assigned to the corresponding total score (Quinsey et al. 2006). If an offender's total score falls near the cusp of a risk category, a range may be provided to reflect it (Quinsey et al. 2006).

The authors of the VRAG contend that this particular instrument is far superior to clinical judgment, although they do not define precisely what clinical judgment means. Moreover, they are somewhat inconsistent in their opinion about the role of clinical judgment. After developing the Violence Prediction Scheme, Webster et al. (1994) acknowledged that predictive accuracy may be improved if modest adjustments are made, but *only* if the evaluator has good reason to believe that a particular factor is related to the likelihood of violent recidivism in a respondent (p. 57). Four out of the five Violence Prediction Scheme developers crafted the VRAG shortly after and, in their description, offered a different view: "Actuarial methods are too good and clinical judgment is too poor to risk contaminating the former with the latter" (Quinsey et al. 1998, p. 171). Hart (1998) observed that this is an extreme position and that such exclusive reliance on actuarial methods is not appropriate.

It is important to clarify that actuarial assessments, such as the VRAG, will tell us if an individual belongs to a group likely to commit violence; however, they are not able to distinguish which specific members of the said group will definitely commit violence. Actuarials also do not take into account any rare risk or protective factors. It is quite a different approach from the MacArthur studies which we have described, in which they use actuarial assessment instruments as tools for clinical assessment.

Sex Offender Risk Appraisal Guide (SORAG)

Initially developed as an actuarial measure for determining risk of sexual recidivism (Quinsey et al. 1995), the SORAG has been expanded to measure for the risk of *any* type of violent re-offending in adult male sexual offenders. Violent re-offending, as defined by the developers, refers to both sexual and nonsexual forms of violence that involve physical contact with a victim and (Quinsey et al. 1998, 2006). The instrument was not designed to be used with females or with offenders younger than 18 years of age.

The SORAG is a modification of the VRAG that consists of 14 weighted items, 10 of which were retained from the VRAG. The SORAG's development was also retrospective in its design and utilized data collected from a series of follow-up investigations using samples of sexual offenders referred to a forensic psychiatric facility in Canada (Quinsey et al. 1995). Overall, the risk factors were chosen based on whether or not they provided unique contributions to the prediction of violent recidivism. This includes risk factors such as deviant sexual preferences and history of sexual offenses (Quinsey et al. 2006). The SORAG's method was conceptualized in the same two books as the VRAG.

As was mentioned earlier, the SORAG shares 10 items in common with the VRAG, specifically: (1) Separation from parents before age 16; (2) Elementary school maladjustment; (3) History of alcohol problems; (4) Marital status; (5) Criminal history score for nonviolent offenses prior to the index offense; (9) Failure on prior conditional release; (10) Age at index offense; (11) Meets DSM criteria for any personality disorder; (12) Meets DSM criteria for schizophrenia; and (14a) Psychopathy (PCL-R Score) or (14b) Psychopathy (CATS Score) (Quinsey et al. 2006).

The items for which there is no overlap with the VRAG include: (6) Criminal history for violent offenses; (7) Number of previous convictions for sexual offenses; (8) History of sex offenses only against girls under age 14, including the index offense; and (13) Phallometric test results (Quinsey et al. 2006). The incorporation of phallometric test is useful for detecting the presence of sexual deviancy (Campbell 2007).

For the purposes of coding, a variety of information should be obtained to maximize accuracy. This includes the facts pertaining to the index offense(s), as well as details concerning the offender's conduct and behavior as a youth, family and relationship history, history of criminal activities and antisocial behavior, and psychological diagnoses (Quinsey et al. 2006). While interviewing the offender is not required, useful information could probably be obtained through such interaction (that is, certain behavioral traits may be observed that corroborate outside information). After the requisite information has been gathered, the scoring may commence. Like the VRAG, once again, a coding sheet is provided to score the measure. To begin scoring, the evaluator simply goes through the possible scoring options and selects the response most supported by the available information.

Total SORAG scores range from −25 to +51 and, like the VRAG, they correspond to 7- and 10-year probability estimates for violent recidivism. Further, the total score is also used in conjunction with a conversion table that is provided in the instrument's protocol for interpreting the results. This table is divided into two columns: SORAG score and Category of Risk, which denotes the risk level assigned to the corresponding total score (Quinsey et al. 2006). To determine the risk level, the evaluator simply scrolls down the SORAG table until the offender's total score is located. After the offender's total score has been located, the evaluator then records the risk category designated for the respondent's respective total score. If an offender's total SORAG score falls near the cusp of a risk category, a range may be provided to reflect the elevated risk (Quinsey et al. 2006).

Domestic Violence Risk Appraisal Guide (DVRAG)

The DVRAG is a 14-item actuarial risk assessment tool designed to identify the level of risk posed by male perpetrators of intimate partner violence (IPV) (Hilton et al. 2008). The DVRAG is part of a system that includes the Ontario Domestic Assault Risk Assessment (ODARA) (Hilton et al. 2004), a shorter version of the DVRAG to be used by frontline officers to inform decision-making, such as bail (Hilton et al. 2008). The DVRAG, on the other hand, provides a more comprehensive assessment (Hilton et al. 2010). The DVRAG includes all 13 of the ODARA items plus a score on the PCL-R.

Construction of an actuarial measure to determine risk of partner assault, specifically, was largely inspired by the success of the VRAG for gauging the risk of future wife assault in perpetrators of IPV (Hilton et al. 2001). In fact, it was entitled the DVRAG to reflect its connection with the VRAG (Hilton et al. 2008) and it shares many points to consider during the coding phase with its predecessors. For example, all of the instruments include measures of clinical history and the PCL-R (Hilton et al. 2010). The DVRAG demonstrated an early ability to discriminate between groups of high- and low-risk offenders during its standardization (Hilton et al. 2010). As noted by Hilton et al. (2010), the specific items include:

(1) *Number of prior domestic violence incidents*: This item covers all prior instances of physical assault by the perpetrator against a current/previous female cohabiting partner or her children. The incident must have occurred on a separate date before the index assault and must be recorded in a police report or criminal record;

(2) *Number of prior nondomestic incidents*: This includes physical assaults by the perpetrator against victims other than the current or previous female cohabiting partner or her children that occurred prior to the index assault;

(3) *Prior custodial sentence of 30 days or more*: This is scored if the offender has ever been sanctioned and incarcerated for at least 30 days prior to the index assault;

(4) *Failure on prior Conditional Release*: This item should be scored "No" if the offender was *not* on Conditional Release at the time of the index assault and/or has no history of failing to comply with Conditional Release conditions;

(5) *Threats to harm or kill at the index assault*: This item is scored if the perpetrator made threats to harm others during the time of the index assault. It may also be scored if the perpetrator was charged with threatening or an equivalent offense;

(6) *Confinement of the partner at the index assault*: This item should be scored positively if the perpetrator confined the victim to a locked area, a barricaded area, or physically prevented the victim from leaving during the index assault;

(7) *Victim concern*: This variable should be scored positive if the female victim of the index assault verbally stated that she was concerned, afraid, worried, or certain that the perpetrator would assault her or her children in the future;

(8) *Number of children*: This includes the number of living minor or adult biological or adopted children of the perpetrator, as well as the additional children of the victim;

(9) *Victim's biological children from a previous partner*: This item pertains to all the living minor or adult biological children of the female victim of the index assault, whose father is not the perpetrator. Adopted children should not be included;

(10) *Violence against others*: This is scored if the perpetrator has any nondomestic physical assaults prior to the index assault. The incident does not need to be officially recorded in a police report or Court documents;

(11) *Substance abuse score*: This variable is based on the number of factors present that indicate substance abuse concerns;

(12) *Assault on victim when pregnant*: This item is scored regardless of whether or not the incident is documented in a police report;

(13) *Number of barriers to victim support*: This item refers to the number of barriers present that interfere with the victim's access to support services; and,

(14) *Psychopathy Checklist-Revised* (*PCL-R*) *score*: As in other cases, a previously documented PCL-R score may be used if there is inadequate information available.

The evaluator scores the DVRAG according to a coding sheet that is provided. To begin with, the examiner goes through the list of possible scoring options and picks the ones most supported by the available information. The item scores are then tallied for a final total score. The possible range of scores on the DVRAG ranges from −10 to +46. The total score is translated using a table that is provided with the measure that indicates the perpetrator's probability of recidivism. The score and likelihood of recidivism can be used for safety planning, as well as for determining the intervention and resources necessary to prevent future violence (Hilton et al. 2010). Higher scores generally reflect greater risk, although the DVRAG cannot actually provide estimates regarding the likelihood of deadly assaults. However, the presence of additional factors may indicate increased reason for concern, such as use of firearms, the termination of the relationship, and/or the victim having a new partner (Hilton et al. 2010).

Static-99

The Static-99—so named to reflect the tool's rather singular focus on "static" variables—is a 10-item actuarial risk assessment instrument crafted to determine the level of risk for sexual and/or violent re-offending posed by adult male sexual offenders (Hanson and Thornton 1999). The offender must have already been convicted of at least one sexual offense against a non-consenting adult or child (Campbell 2007). It should not be used with offenders younger than 18 years of age at the time of release or with female offenders. The normative data used to interpret a respondent's score was updated in 2009 to include a wider variety of sex offenders and increase the instrument's generalizability (Helmus et al. 2009). Indeed, the Static-99 has been described by some as the most widely used (McGrath et al. 2010) and most empirically studied risk assessment instrument for use with adult male sexual offenders (Hanson and Morton-Bourgon 2009).

The Static-99 is composed of 10 items, which are scored using the 2003 revised coding rules (Harris et al. 2003). The measure solicits information from three areas in an offender's life: Demographic information, criminal history, and victim information (Harris et al. 2003, p. 11). Each of the items is listed with three elements attached, namely "The Basic Principle," "The Basic Rule," and a list of criteria to score the item (Harris et al. 2003, p. 1). Scores range from 0 to 12 and are used to determine an offender's placement into one of four risk categories, which reflects the respondent's long-term risk potential.

As previously indicated, the Static-99 is divided into three dimensions. The first, Demographic Information, is quite straightforward and includes the following items:

(1) *Youth*: This item asks if the offender will be between his 18th and 25th birthday at exposure to risk—In other words, will the offender be at least 18 years of age when released and, subsequently, given the chance to re-offend? Justification for inclusion of the item is based on research (Hanson 2001) reflecting that a younger age is related to an increased risk of recidivism. If the evaluation is looking at the potential for re-offending at a future point in time, the age at which he will be exposed to risk is used (Hanson and Thornton 2003); and,

(2) *Ever lived with an intimate partner—Two years*?: This item inquires whether or not the offender has ever lived with an intimate partner for at least 2 years. This item was selected based on Hanson and Bussiere's (1998) meta-analysis concerning risk factors for sexual recidivism. Ultimately, it is a reflection of whether the offender had the personality and resources needed to maintain a relatively stable relationship with another adult, regardless of whether it a heterosexual or homosexual relationship (Harris et al. 2003). The examiner must make sure that the quality of the relationship meets the criteria for the item, which includes a list of factors to be considered (e.g., extended absences). This is the only item on the tool that can be omitted if there is insufficient information available.

The second dimension on the Static-99 inquires about a number of factors related to criminal history. This section comes with a very comprehensive list of rules and exclusionary criteria for coding that should be reviewed prior to rating the assessment, especially since different interpretations of the items can lead to different outcomes. For example, it is possible for an offender to commit a number of sexual offenses in different jurisdictions over a protracted period—be it during a spree or spread out over many years—before finally being detected or arrested. When such is the case, the grouping of charges and convictions would constitute an "index cluster" rather than multiple index "offenses" (Harris et al. 2003). This is because the offender was not caught, sanctioned, and then chose to re-offend yet again after being released. The criminal history section includes the following items:

(3) *Index Nonsexual Violence—Any Convictions*?: This items asks if there was a conviction for a separate, nonsexually violent offense addressed on the same sentencing occasion as the index sex offense. It is based on research (Hanson and Bussiere 1998) indicating that a history of nonsexual violence is predictive of future violence, as well as the potential for serious injury if the individual does recidivate (Harris et al. 2003);

(4) *Prior Nonsexual Violence—Any Convictions*?: This item is similar to the previous item, but refers to any conviction that is not part of the index offense. Also like item #3, this item was included based on research (Hanson and Bussiere 1998) indicating that a history of nonsexual violence is, again, predictive of future violence and the potential for serious injury (Harris et al. 2003). It may also reflect difficulties with impulse control and an unwillingness (or inability to) conform their behavior, even after being disciplined;

(5) *Prior Sex Offenses*?: This item is used to identify patterns that reflect a higher likelihood of future re-offending. It includes both *prior* charges and convictions, which are summed separately; charges that were dropped or did not result in a conviction are also included. After the number of charges and convictions has been totaled, the raw scores are converted to final scores. The highest final score is the score that is recorded. The index offense is not included when tallying the final score for this item (Harris et al. 2003);

(6) *Prior Sentencing Dates*?: This item seeks to establish a pattern of criminal behavior by asking if the offender has either had four or more or less than four sentencing dates prior to the index offense (Harris et al. 2003). Ultimately, it is asking if the offender can be deterred by a fear of the consequences; and,

(7) *Any Convictions for Noncontact Sex Offenses*?: This item involves offenses such as voyeurism and exhibitionism. It was included based on the rationale that most individuals have little interest in exposing themselves to strangers, stealing underwear, etc. (Harris et al. 2003). The authors also note that offenders who engage in these types of behaviors are more likely to have problems altering their behavior to match conventional standards (Harris et al. 2003). Ultimately, it reflects the likelihood of the offender modifying his maladaptive sexual behavior.

The final portion of the Static-99 is referred to as "The Three Victim Questions." Not surprisingly, this section addresses the offender's choice of victims. It is important to note that these items do not apply to nonsexual offenses or sexual offenses relating to prostitution, offenses against animals, possession of child pornography, and public sex with consenting adults (Harris et al. 2003). The items for this section include:

(8) *Any Unrelated Victims*?: The manual cites research (i.e., Harris and Hanson, unpublished manuscript) demonstrating that offenders who only aggress against those who are related to them are less likely to re-offend than individuals who offend against those they are not related to. A victim is considered to be related if the relationship is sufficiently close that marriage would be prohibited. Additional criteria needed to determine relatedness are provided in the coding manual (Harris et al. 2003);

(9) *Any Stranger Victims*?: This item (Harris et al. 2003) is very similar to the one preceding it, but it more specifically reflects whether the offender is driven enough to seek out a target who cannot identify him so that he can avoid detection; and,

(10) *Any Male Victims*?: This item is based on research suggesting that offenders who victimize males are more likely to re-offend in the future than offenders who exclusively victimize females (Hanson and Bussiere 1998). This includes non-consenting adults and children.

In terms of coding, the evaluator should rely on information that can be corroborated by outside sources. Scoring the items based on self-report is strongly discouraged, yet some flexibility is allowed. For example, if it is impossible to obtain file information or access collateral sources, some limited usage of self-report is acceptable; however, the information needs to seem credible and reasonable to the evaluator (Harris et al. 2003). The specific information needed to score the Static-99 depends on which of the three categories is being assessed (Harris et al. 2003). According to the coding manual, the demographic items can be scored using information derived from official records or collateral sources. Regarding the offender's criminal history, the evaluator must have access to the official criminal record. For the items addressing an offender's victim selection, the use of all credible information is recommended, including self-report, victim and/or witness statements, and third-party sources.

The evaluator may begin scoring the Static-99 after the requisite information has been gathered. This phase is completed on a coding sheet that is provided. As with the previously mentioned instruments, a rating is given after the evaluator has looked through the response options and selected that which is most supported by the available information. To determine the total score for the measure, the evaluator simply sums the scores for each of the individual items. Again, total scores can range from 0 to 12 and reflect estimates of future risk based on the number of risk factors present in any one individual (Harris et al. 2003). In general, the more risk factors the offender endorses, the more likely he is to recidivate in the future (e.g., Austin et al. 2003).

After the total score for the Static-99 has been calculated, the evaluator consults the table entitled, "Static-99 Recidivism Percentages by Risk Level," which

includes recidivism risk estimates for both sexual and violent recidivism over 5-, 10-, and 15-year periods (Harris et al. 2003). The evaluator notes the specific recidivism rates (the more recent norms should be used for more geographically diverse populations). Although the tool is useful for estimating long-term risk potential (Conroy and Murrie 2007), it is not possible to determine specifically when he will recidivate.

As was the case with the previously discussed actuarial measures, the Static-99's development was retrospective in design and based on facts readily available in database records. The risk factors were first identified during a review of the empirical literature, after which statistical analyses were performed to determine which risk factors made the greatest contributions (Hanson and Thornton 1999).

A modified version of the Static-99 was later developed and, not surprisingly, it was entitled the Static-2002 (Hanson and Thornton 2003). This measure is unique from the Static-99 because it includes some theoretically meaningful variables thought to drive sexual offender recidivism (Phenix et al. 2008). Both measures were then revised, resulting in the development of the Static-99R and Static-2002R (Helmus et al. 2011). Both instruments are essentially identical to the predecessors, but with one major modification: The effect of aging on recidivism is taken into account to minimize overestimation of recidivism risk in older offenders. Thus, the item addressing an offender's age at release is weighted differently and, therefore, each response option is assigned a different point value than that which is assigned in the original instruments (Helmus et al. 2012). This is the only item for which the effect of aging is reflected.

Coding the Static-99R and Static-2002R entails the same process that was used with the other instruments. Basically, the rating is completed on a coding sheet that is provided, the most supported option is selected, and the total score is simply a summation of all individual item scores. On the Static-99R, 0 to 1 is considered Low risk; 2 to 3 is deemed a Moderate-Low risk; 4 to 5 is listed as a Moderate-High risk; and values of 6 or greater are regarded as High risk (Helmus et al. 2011). On the Static-2002R, a score that is less than or equal to 2 is classified as Low risk; a score of 3 or 4 reflects a Moderate-Low risk; a score of 5 or 6 indicates Moderate risk; a score of 7 or 8 suggests Moderate-High risk; and a value of 9 or greater is regarded as High risk (Helmus et al. 2011). Hanson et al. (2012) made available a systematic method for relaying these results, namely percentile ranks, to minimize ambiguity.

Rapid Risk Assessment for Sex Offender Recidivism (RRASOR)

The RRASOR is an actuarial risk assessment instrument designed to determine the level of risk posed by adult males who have already been convicted of at least one sexual offense (Hanson 1997). It was designed to be a brief instrument for

identifying offenders who are at high risk for sexual re-offending. Indeed, this particular measure is quite useful for settings that entail routine screening of sex offenders. Hanson (1997) has noted, though, that it should only be used as a screening instrument to place offenders into relative risk levels until a more comprehensive assessment can be performed. Thus, the RRASOR should never be used in place of a more comprehensive measure. Like the other tools described thus far, the measure was not crafted for use with offenders who are less than 18 years of age or female offenders.

As previously indicated, there are only four items on the RRASOR (each of which is also found on the Static-99). These risk factors were selected from Hanson and Bussiere's (1996) meta-analysis concerning risk markers for sexual recidivism. Namely, the instrument includes: (1) Prior Sexual Offenses; (2) Being Between the Ages of 18 and 25 years Old; (3) Victim Gender; and (4) Any Unrelated Victims.

Coding the RRASOR is fairly straightforward. The first item, for instance, necessitates a simple summation based on the number of arrests or the number of convictions in an offender's criminal record. Like the Static-99, whichever option produces the higher score is selected for the final tally. It is important to note that this item is the most heavily weighted item on the RRASOR (Hanson 1997), which is not surprising as the best predictor of future behavior is past behavior. Only official records should be used to code this item. The second item, which addresses the offender's age, is based on the time of first opportunity to offend. The third item, victim gender, requires a review of all available information, including the offender's self-report, official records, collateral sources, and case notes (Hanson 1997). The final item, relationship to victim, includes spouses and family members who are too closely related to get married. Similarly, if the offender is in a caretaker role in the same house as the victim, they are also considered related (Hanson 1997). The evaluator scores on a provided coding sheet and must, of course, go through each of the possible scoring options. Total scores range from 0 to 6 with 10-year estimates concerning the likelihood of recidivism.

The items evaluated during the instrument's development were drawn from Hanson and Bussiere's (1996) meta-analysis concerning risk markers for sexual recidivism. An item was identified for consideration if it demonstrated a strong correlation with sexual recidivism and could be scored using commonly available information (Hanson 1997). Items were then either ruled-out or selected for inclusion in the final instrument based on the results of regression analyses using six samples of sexual offenders (Hanson 1997). The four items selected were identified as the best combination to most accurately predict the outcome of sexual recidivism in a brief amount of time; The items were then weighted according to their ability to predict the likelihood of recidivism over periods of 5–10 years (Hanson 1997).

Minnesota Sex Offender Screening Tool—Revised (MnSOST-R)

The MnSOST-R is an actuarial risk assessment instrument designed to assess the level of risk for sexual re-offending in adult male rapists and extra-familial child molesters who are incarcerated (Epperson et al. 2003). It was not designed for use with juveniles, females, offenders on probation, or intra-familial child molesters. Often, it is regarded as a screening tool for referral for commitment under Minnesota's Sexual Psychopathic Personality and Sexually Dangerous Persons Law (Epperson et al. 2003). The measure covers both static and dynamic variables (this is the only one of the actuarial instruments that makes specific reference to "dynamic" variables). Its predecessor, the Minnesota Sex Offender Screening Tool, had some clinical observations built in but the current revised version has variables based only on actuarial data (Conroy and Murrie 2007).

The MnSOST-R consists of a total of 16 items. The first 12 items are regarded as "Historical/Static variables" and the last four are considered the "Institutional/Dynamic Variables." The items identified for review by Epperson et al. (2003) are as follows:

(1) Number of sex/sex-related convictions (including current conviction);
(2) Length of sexual offending history;
(3) Was the offender under any form of supervision when s/he committed any sex offense for which they was eventually charged of convicted?;
(4) Was any sex offense (charged or convicted) committed in a public place?;
(5) Was force or the threat of force ever used to achieve compliance in any sex offense (charged or convicted)?;
(6) Did any sex offense (charged or convicted) involve multiple acts on a single victim within any single contact event?;
(7) Number of different age groups victimized across all sex/sex-related offenses (charged or convicted). The age groups of victims include 6 and younger; 7–12; 13–15, and the offender is more than 5 years older than the victim; and 16 and older;
(8) Offended against a 13–15-year-old victim and the offender was more than 5 years older than the victim at the time of the offense (charged or convicted);
(9) Was the victim a stranger in any sex/sex-related offense (charged or convicted);
(10) Is there evidence of adolescent antisocial behavior in the file?;
(11) Pattern of substantial drug or alcohol abuse (12 months prior to arrest for instant offense or revocation); and,
(12) Employment history (12 months prior to the arrest for instant offense or revocation).
(13) Discipline history while incarcerated;
(14) Chemical dependency treatment while incarcerated;
(15) Sex offender treatment history while incarcerated; and,
(16) Age of offender at time of release.

The scoring instructions, which were revised by Epperson et al. (2005), direct the evaluator to review the offender's file prior to assessment. Only information documented in official data sources should be used to score the instrument. Self-report information may be included on certain items (i.e., items 2, 10, 11 and 12), but only if it is already documented and determined to be reliable (Epperson et al. 2005). Offenses that were charged but not convicted are included if there is evidence that a sex offense was, in fact, attempted or perpetrated (Epperson et al. 2005).

The MnSOST-R was developed with a focus on making the instrument relatively easy to score to ensure that correctional personnel could use it effectively (Hoberman 2001). Scoring the MnSOST-R is fairly straightforward, in that the evaluator is tasked with identifying the response option best supported by the available information, selecting said item, and eventually tallying all item scores together to produce a total score (Epperson et al. 2003). As far as interpretation, an offender can be placed into one of three risk categories based on his score. Namely, a score of 3 or below is designated as Low (or Level 1); 4–7 is classified as Medium (or Level 2); and a score of 8 or above is regarded as High (or Level 3) (Epperson et al. 2003).

The items on the MnSOST-R were derived using information available in the Minnesota Department of Corrections offender files, which have been described as more comprehensive and detailed compared to the file information readily available in other states (e.g., DeClue 2002). As a result, the measure's inter-rater reliability has been described as very low because information necessary to score some of the items may not be available in an offender's file if he resides in a state other than Minnesota (Austin et al. 2003). After an item was selected for inclusion in the final instrument, each answer was weighted in accordance with its purported ability to predict sexual recidivism (Epperson et al. 2003). As noted before, the revised version was derived from empirical methods rather than clinical observation to ensure improved predictive validity (Epperson et al. 2003).

Structured Professional Judgment

The instruments to be described in this section are not actuarial, at least in the sense that the previously described tools were. Rather, they utilize structured professional judgment (SPJ) methodology. Structured Professional Judgment—or, as it is sometimes called, structured clinical judgment—attempts to combine all of the psychological research regarding violent or sexually violent offending into a series of distinct categories, and then uses a structured interview format to ensure all areas are adequately covered. Therefore, unlike the actuarial approach, there is a good deal of leeway for clinical judgment and clinical interview; However, the areas covered are very carefully derived from the available research. Some authors regard it as an amalgamation of the actuarial and clinical methods.

Historical, Clinical, Risk Management-20 (HCR-20)

The first instrument that we will discuss within this category is the Historical, Clinical, Risk Management-20 (HCR-20) scheme, Version 2 (Webster et al. 1997). The HCR-20 is a SPJ scheme "intended to be used to guide a comprehensive, structured assessment of violence risk for adults (18+) within forensic psychiatric, civil psychiatric and offender samples" (Douglas et al. 2014). This measure is basically a formalized guide for the systematic assessment and management of violence risk. (Conroy and Murrie 2007) and represents one of the first attempts to join clinical methods of assessment with the available research on the prediction of violence (Webster et al. 1997). For the purposes of the HCR-20, violence is defined as "actual, attempted, or threatened harm to a person of persons" (Webster et al. 1997).

The HCR-20 is comprised of 20 items that address historical (H), clinical (C), and risk management (R) factors. It considers dynamic risk factors and, in that way, differs from most of the actuarial assessment instruments. Unlike the MnSOST-R, which is the only actuarial that makes reference to "dynamic" variables, these factors are not limited to behavior in an institution. Such an approach is useful for monitoring change in risk status over time and, for this reason, it can be useful for case management and treatment planning (Webster et al. 1997).

The HCR-20 assesses the risk for future violence by providing a review of three specific areas in an individual's life: his/her past and present behavior, as well as potential future circumstances. Past behavior, specifically, is captured by the Historical scale. The items in this section are often given the greatest weight because they are, technically, actuarial risk factors (Borum 2007). In other words, these items are listed on many of the actuarial measures and their utility has already been discussed in the literature. Despite the overlap, there is more of a "clinical feel" to these items than in the actuarial assessments. The Clinical scale, on the other hand, is far more individually oriented, as well as more suited for determining short term-risk of violence (e.g., inpatient violence; McNeil et al. 2003) and treatment implications. Finally, situational factors that can occur post-treatment are found on the Risk Management scale. These items prompt a proactive evaluation of future circumstances to minimize risk of relapse/decompensation. The manual notes that it may be helpful to consult with those responsible for treatment or community release when completing this section (Webster et al. 1997). The specific HCR-20 items are as follows:
Historical:

(H1) Previous violence;
(H2) Young age at first violent incident;
(H3) Relationship instability;
(H4) Employment problems;
(H5) Substance use problems;
(H6) Major mental illness;
(H7) Psychopathy;

(H8) Early maladjustment;
(H9) Personality disorder; and,
(H10) Prior supervision failure.

Clinical:

(C1) Lack of insight;
(C2) Negative attitudes;
(C3) Active symptoms of major mental illness;
(C4) Impulsivity; and,
(C5) Lack of responsiveness to treatment.

Risk Management:

(R1) Future plans lack feasibility;
(R2) Exposure to destabilizers;
(R3) Lack of personal support;
(R4) Noncompliance with remediation attempts; and,
(R5) Stress.

The coding process for the HCR-20 was "cloned from the PCL-R" (Cawood and Corcoran 2009), meaning it utilizes a similar three-point system. Specifically, a score of 0 is assigned if the risk factor is not present or is contraindicated by the available information; a 1 if the available information suggests the possible presence or mild presence of an item; and a score of 2 if the available information definitely indicates the item's presence (Webster et al. 1997). If there is insufficient information to code an item, the evaluator is given the option to omit that item (Arbisi 2003). Each item is concretely defined in the manual to provide clarity. As in the PCL-R, multiple sources of information are needed in order to adequately assess the presence of each risk factor.

Scoring the HCR-20 is completed on a coding sheet provided. According to Webster et al. (1997), there is little sense in using random fixed cutoff scores to determine an individual's placement into a risk category. Rather than merely summing the scores (as was done with the actuarials), it is recommended that the evaluator use the information to assign the examinee to a risk classification level—Low, Moderate, or High—based on the number of risk factors present in the specific individual being assessed (Webster et al. 1997). Also, in contrast with actuarial methodology, reliance on group-based scores is discouraged.

Again, the HCR-20 is an example of structured professional judgment, rather than an actuarial instrument. It was designed to establish whether or not certain important risk factors are present in a respondent (like, a checklist) rather than produce a numerical summation. Thus, it is recommended that the results be listed as risk and protective factors rather than numbers because numerical results could be misleading (Webster et al. 1997). The authors further note that there are certain items that, on their own, could render and individual at high risk for future violence. For example, the items concerning substance abuse, psychopathy, poor insight, difficulty inhibiting one's behavior, and noncompliance with remediation could be ominous indicators. Again, it is obvious that clinical judgment plays a significant role in interpreting these various factors.

Sexual Violence Risk-20 (SVR-20)

The SVR-20 is an example of structured professional judgment designed to identify individuals at risk of committing sexual violence (Boer et al. 1997). Like the HCR-20, this instrument assists in the systematic review of critical factors (many of which are shared by the two instruments); however, the SVR-20 also contains certain risk factors that are more specific to sexual offending. It also allows the individual examiner to, once again, pursue each of the variables using their own clinical judgment.

As noted above, the SVR-20 is a list of structured professional judgment guidelines that highlight variables in an offender's life to explore during a thorough assessment. The SVR-20 includes items that address three areas in an offender's life: Psychosocial adjustment, sexual offenses, and future plans (Boer et al. 1997). The SVR-20 also includes psychophysiological data, such as penile plethysmographic evaluation. The specific items include:

A. Psychosocial adjustment:

    (1) Sexual deviation;
    (2) Victim of child abuse;
    (3) Psychopathy;
    (4) Major mental illness;
    (5) Substance abuse problems;
    (6) Suicidal/homicidal ideation;
    (7) Relationship problems;
    (8) Employment problems;
    (9) Past nonsexual violent offenses;
  (10) Past nonviolent offenses; and,
  (11) Past supervision failure.

B. Sexual offenses:

  (12) High-density sex offenses;
  (13) Multiple sex offense types;
  (14) Physical harm to victim;
  (15) Use of weapons or threats of death in sex offenses;
  (16) Escalation in frequency or severity of sex offenses;
  (17) Extreme minimization or denials of sex offenses; and,
  (18) Attitudes that support or condone sex offenses.

C. Future Plans:

  (19) Lacks realistic plans; and,
  (20) Negative attitude toward intervention.

There is an additional section on the SVR-20, entitled "Other Considerations," that does not include any items to be coded. Rather, it may be used to document unique risk factors and idiographic data that may have influenced the final judgment regarding level of risk (Boer et al. 1997).

Barbaree et al. (2008) reported that including information about an offender's age-at-release can improve the measure's predictive accuracy. More specifically, they found that 13 out of the 20 items on the SVR-20 reflect a general anti-social lifestyle and, when age was corrected for, there was a subsequent reduction each item's predictive power. These items include: (2) Victim of child abuse; (3) Psychopathy; (5) Substance abuse problems; (6) Suicidal or homicidal ideation; (7) Relationship problems; (8) Employment problems; (9) Past nonsexual violent offenses; (10) Past nonviolent offenses; (11) Past supervision failure; (14) Physical harm to victims; (15) Use of weapons or threats of death in sex offenses; (19) Lack of realistic plans; and (20) Negative attitude toward intervention. In short, increasing age diminishes the power of these variables to predict recidivism.

Scoring the SVR-20 is very similar to the process used with the HCR-20: 0 (factor not present), 1 (partially or possibly present), and 2 (definitely present) (Boer et al. 1997). Also like the HCR-20, all available information—including interviews, criminal records, prior mental health reports, the results of standardized psychological measures, and psychophysiological data—should be used when rating the items (Budrionis 2003). If the available information indicates that a risk factor is not applicable to the offender, the examiner moves on to the next item in the list.

As noted in the discussion of the HCR-20, a summation of the individual item scores is not optimal for interpretation. Rather, the SVR-20 should be used as a set of guidelines to determine which information from an offender's history should be considered in order to assign the examinee to a risk classification level of Low, Moderate, or High (Boer et al. 1997).

Spousal Assault Risk Assessment Guide (SARA)

The SARA is a structured professional judgment scheme developed as a clinical guide to evaluate the level of risk posed by perpetrators of intimate partner violence toward partners or ex-partners, children, other family members, and other possible targets (Kropp et al. 1999). The SARA also provides a structured way of reviewing important information and is useful for violence prevention and correctional decisions (Bourgon 2011); it can also be used to determine case management strategies and interventions (Bowen 2011). The 20 items function as a sort of checklist that results in two overall risk estimates: imminent risk of violence toward the partner or ex-partner and imminent risk of violence toward others. However, there is no clear definition of what constitutes "imminent" (Helmus and Bourgon 2011).

The 20 items on the SARA address both static and dynamic risk factors and are broken down into four sections: criminal history, psychosocial adjustment, spousal assault history and current/most recent offense (index offense) (Kropp et al. 1999). There is a final section, "Other Considerations," where idiographic

risk factors that contributed to the overall judgment can be noted (e.g., stalking). The items are as follows:

A. Criminal History:

    (1) Past assault of family members;
    (2) Past assault of strangers or acquaintances; and,
    (3) Past violation of Conditional Release or community supervision.

B. Psychosocial Adjustment:

    (4) Recent relationship problems;
    (5) Recent employment problems;
    (6) Victim of and/or witness to family violence as a child or adolescent;
    (7) Recent substance abuse or dependence;
    (8) Recent suicidal or homicidal ideation or intent;
    (9) Recent psychotic or manic symptoms; and,
    (10) Personality disorder with anger, impulsivity or behavioral instability.

C. Spousal Assault History:

    (11) Past physical assault;
    (12) Past sexual assault/sexual jealousy;
    (13) Past use of weapons and/or credible threats of death;
    (14) Recent escalation in frequency or severity of assault;
    (15) Past violation of No Contact Orders;
    (16) Extreme minimization or denial of spousal assault history; and,
    (17) Items that support or condone spousal assault.

D. Current/most recent offenses:

    (18) Severe and/or sexual assault;
    (19) Use of weapons and/or credible threats of death; and,
    (20) Violation of No Contact Order.

The scoring process for the SARA is similar to that used for the HCR-20 and SVR-20. Thus, a score of 0, 1, or 2 is awarded for each item based on its applicability (Kropp et al. 1999). After the SARA has been scored and the presence of critical items noted, the evaluator makes a final determination of whether the perpetrator is Low, Moderate, or High risk. This segment is referred to as the "summary risk judgment" and results in two possible scores. The first score covers imminent risk of harm to the spouse and the second score addresses imminent risk of harm to another identifiable person (Kropp et al. 1999). As with the other SPJ tools, a simple summation of the item values is discouraged.

The SARA was one of the first measures designed specifically for determining risk of IPV (Hilton et al. 2010), and is frequently described as the most well-research measure for interpersonal violence (Hilton and Harris 2007). Development of the SARA began with a review of the literature to identify traits and behavior patterns that could discriminate between IPV perpetrators who

would re-offend from those who would not (Kropp et al. 1995). Clinical and legal factors were also considered during the item selection process (Kropp and Hart 2000). The normative data were obtained using two groups of adult male offenders, namely offenders on probation and inmates. While all of these subjects had a history of interpersonal violence, not all were incarcerated or on probation for that reason (Katz 2003).

Observations and Limitations

Risk assessment measures, whether actuarial in nature or relying on structured professional judgment, typically have one major element in common: They produce results that are only slightly better than chance. This level seems to hold regardless of whether we are talking about sexual recidivism or violent recidivism Truly, it is not uncommon for studies concerning the predictive validity of risk assessment tools to produce results that conflict with the findings of other studies. Clearly, this is an exceedingly complex area and within this brief overview we cannot discuss these in any great detail. However, a few general comments are in order.

Sampling error. First, let us look at the issue of sampling error. Sampling error is a term used in statistics to refer to discrepancies between the characteristics of a sample chosen to represent a population, as well as when findings are generalized despite it being inappropriate to do so. When we look at the predictive ability of risk assessment instruments, sampling error frequently comes into play and occurs when different groups of sexual offenders or violent offenders are assumed to represent all such offenders. For example, a study that utilizes a sample composed of rapists to determine the predictive validity of a risk assessment instrument for sex offenders is unlikely to produce the same results as would be obtained from a sample of child molesters. This is because the risk factors that predict recidivism in rapists are often distinct from the risk factors that predict sexual re-offending in child molesters (Firestone et al. 2000; Hanson and Morton-Bourgon 2005). Given that these instruments are an amalgam of risk factors, some of which may be more applicable for one sex offender subtype than other, it is not surprising that they have only moderate predictive validity, at best. It may be wiser to construct assessment instruments that measure for specific sex offender subtypes than to continue relying on a "one size fits all" approach.

Even the kinds of predictions made may vary. Becker (2014), in a presentation at the APA-ABA conference on Violence and the Family, noted that certain risk factors may predict violence in general, while others may predict an initial sexual offense, while still others may predict sexual recidivism; According to Becker, far more research needs to be done to "fine-tune" the assessments and exactly what they do predict. Becker further notes that no one etiological theory is widely accepted. Some theories include, but are not limited to, brain structure, hormonal abnormalities, cognitive distortions, distorted social learning and conditioning, and a variety of personality variables. She notes that Antisocial Personality Disorder *is*

a risk factor for those who will rape, but not for those who engage in child moles-tation. In short, according to Becker, sex offenders are a heterogeneous group, have no distinct profile, and come from all walks of life. She suggests that future research follow a model that widens the range of both predictor and criterion vari-ables being studied. Becker further suggests that future research needs to include biological predispositions, family and peer group influences, and the individual's capacity for self-regulation. She notes, for instance, that not all child molest-ers or rapists have a paraphilic disorder, nor do all diagnosed with a paraphilic disorder have offense histories. Incest offenders tend to have fewer victims, and extra-familial sex offenders are more likely to have deviant sexual preferences. Noncontact offenders, such as voyeurs, exhibitionists, and Internet offenders, tend to show more elements of compulsivity and a larger number of victims.

Treatment versus no prior treatment. Another area has to do with offend-ers with a history of treatment, as opposed to those with no prior treatment. An offender's history of treatment may pose additional issues. There are many studies that have used offenders with a history of treatment or currently in treatment while others used offenders with no history of treatment. Clearly, some factors may apply in one group that would not apply in another, and certainly in other areas where there is a purported use of so-called "dynamic variables," the nature of their participation in treatment is not even included. Also, the nature of the treatment needs to be examined as the different orientations encourage different approaches.

Ethnic differences. An offender's ethnicity has also been found to influence the outcome of a study examining an instrument's predictive validity. For instance, Varela et al. (2013) found that the Static-99 (and Static-99R) were far less effec-tive in predicting sexual re-offending with Latino offenders than they were with African-Americans and Caucasians. It may well be premature to assume that these tools are equally predictive across the ethnicities. A frequent occurrence, as observed in court by one of the authors (D.S.), is the tendency of forensic examin-ers to either be oblivious to these findings or just dismissive of them.

Location. The country in which the study took place has also been found to influence the reported predictive ability of an instrument. In a meta-analysis examining the accuracy of assessment measures for identifying an offender's risk of sexual re-offending, Hanson and Morton-Bourgon (2009) determined that the effect sizes produced by such instruments were strongest in the United Kingdom and rather weak in the United States. Since the majority of the assessments were derived in Canada and the United Kingdom, this is not particularly surprising since the United States has a more diverse ethnic make-up. Additional research is needed to determine whether these differences are or are not exclusively attribut-able to ethnic differences.

Research validity versus field validity. Research validity describes how valid a measure is when applied by researchers who have received training in the administration and coding of the assessment. Field validity has to do with the pre-dictive validity of the same measure when administered and scored by clinicians. Often, the research validity surpasses the field validity because the researchers have greater familiarity with the operational definitions and coding criteria used

during a tool's development. Murrie et al. (2012), for instance, found this to be the case in some of their evaluators scoring the PCL-R. Similarly, in one recent study (Singh et al. 2013) it was found that studies authored by an instrument's designers reported predictive validity rates almost twice a high as those found by independent researchers. They identified this as a result of "authorship bias."

Operational definitions of recidivism. Another factor that may relate to the applicability of a study has to do with the definition of recidivism. Some studies may define recidivism as a new arrest, a new charge, or a new conviction. Therefore, it stands to reason that much of the research reportedly evaluating for the same general outcome is really looking at different things. An examiner must keep these details in mind before generalizing the findings of any study.

Actuarial scoring of Structured Professional Judgment measures. A validity issue unique to structured professional judgment studies relates to the way in which the instrument was coded. Structured professional judgment, as noted above, generally produces summary risk ratings rather than absolute actuarial scores. Nevertheless, a good portion of the literature regarding structured professional judgment's validity is based on studies that did not score the instrument according to how it was designed. According to Hanson and Morton-Bourgon (2007), this basically results in the measure being transformed from a structured professional judgment measure to a "conceptual actuarial measure" (p. 3). Therefore, it would be inappropriate to generalize the validity reported by such studies to that which might be found if the instrument were coded as intended, as it introduces sampling error.

Inter-rater reliability. Although structured professional judgment measures were not intended to be actuarially scored, the benefit of such an approach could be improved inter-rater reliability, as summary risk judgments are often based on informed clinical judgment, which may, in fact, vary from one examiner to another. Certain examiners using the SARA, for instance, may give extra weight to stalking behavior endorsed by a perpetrator of interpersonal violence, while a different examiner may not. If such is the case, different summary risk ratings may be expected.

Base rate issues. Going back to the early work on risk assessment, Monahan observed that base rate data in frequently missing in these circumstances and, therefore, any statistics regarding likelihood of offending cannot be truly put within context. The results of an assessment are often adjusted to reflect the number of offenders that recidivate in a particular area. Such estimates may be inaccurate because sexual and violent offense frequently go unreported. In fact, the U.S. Department of Justice (2014), in their annual survey concerning criminal victimization, found that fewer than half of those who experienced violent victimization (46 %) and sexual assault victimizations (35 %) reported it to the police. Overall, low or inaccurate base rates inevitably create issues for the accuracy of violence risk assessment measures (Szmukler 2001).

Group versus individual data. As noted before, one of the major concerns is that the actuarial approach will only point to the fact that the individual belongs to a group that has a particular likelihood of re-offending. It does not say whether

or not that particular individual will re-offend, with the latter estimate needing to be made based on idiographic and clinical data. Therefore, while there are many claims that the actuarial approach is superior to the clinical approach, this really omits the fact that the current approach to risk assessment is primarily that of structured professional or structured clinical judgment, not the unaided clinical judgment that was used previously. It is, in fact, remarkable that textbooks and presentations at meetings still use the artificial distinction between actuarial (the scientific) and unaided clinical (the nonscientific) approach in order to make their assertions. The concept of structured professional judgment as representing an amalgamation of the best in both areas appears to be sorely lacking.

Another area desperately in need of more research is the interactive effect of various risk factors. Dvoskin (2014) in the APA/ABA conference on Violence and the Family, noted the misperception propagated by the media: That mental illness and violence are highly correlated. In fact, only 5 to 10 % of violent crime can be attributed to severe mental illness alone; while most mentally ill individuals do not act violently, they often have other risk factors such as child abuse, victimization, unemployment, and substance abuse; co-occurrence with substance abuse, for instance, is known to multiply the base rate by a factor of 3. Dvoskin also notes a high correlation between those few mentally ill individuals who do act violently and suicidal ideation/behavior on their parts. He recommends doing a suicide risk assessment as part of any violence risk assessment. Clearly, a more complex analysis, such as the MacArthur studies, needs to be utilized on a widespread basis.