

2011

ANALYSIS OF THE DEVELOPMENT AND APPLICATIONS OF RISK ASSESSMENT FOR SEX OFFENDERS IN CORRECTIONAL POPULATIONS: A COMPARATIVE REVIEW

Derek Nicholas Emerson
The University of Montana

Let us know how access to this document benefits you.

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Recommended Citation

Emerson, Derek Nicholas, "ANALYSIS OF THE DEVELOPMENT AND APPLICATIONS OF RISK ASSESSMENT FOR SEX OFFENDERS IN CORRECTIONAL POPULATIONS: A COMPARATIVE REVIEW" (2011). *Graduate Student Theses, Dissertations, & Professional Papers*. 142.
<https://scholarworks.umt.edu/etd/142>

This Professional Paper is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mail.lib.umt.edu.

ANALYSIS OF THE DEVELOPMENT AND APPLICATIONS OF RISK ASSESSMENT FOR SEX OFFENDERS IN
CORRECTIONAL POPULATIONS: A COMPARATIVE REVIEW

By

DEREK NICHOLAS EMERSON

BA Sociology, University of Montana, Missoula, Montana, 2009

Professional Paper

presented in partial fulfillment of the requirements

for the degree of

Master of Arts

in Sociology, Option Criminology

The University of Montana

Missoula, MT

May 2011

Approved by:

J.B. Alexander Ross, Associate Dean of the Graduate School

James Burfeind, PhD.

Sociology

Dusten Hollist PhD.

Sociology

Jeff Greene, PhD.

Political Science

Analysis of the Development and Applications of Risk Assessment for Sex Offenders in Correctional Populations: A Comparative Review

Chairperson: James Burfeind, Ph.D.

Dusten Hollist, Ph.D.

Jeff Greene, Ph.D.

This review discusses the content and theory that drive the creation of risk assessment of sex offenders. While risk assessment tools are used by corrections agencies to determine the potential risk for reoffending for all types of offenders, this paper focuses on those used for sex offenders. The intent with this project is to assess various risk assessment tools and describe the implications on possible new scales, and to cover potential solutions to problems inherent to the field of risk assessment. These goals are accomplished by extensive inventory of the most widely used risk scales, followed by discussion of a series of viable solutions to the problems inherent with risk scales. Suggested changes to the field of risk assessment include incorporating step-wise applications and using theoretically- grounded scale components. This paper also organizes much of the current literature about risk assessment and highlights common shortcomings and innovations of each scale discussed.

The individualized treatment models used in the system of American corrections today present the public with some major dilemmas once treatment programs start dealing with sex offenders. Sexual offenders make up an offender group that is of weighty public concerns, and in need of corrective treatment. One important facet of the treatment process is the task of determining which sex offenders are at greatest risk for reoffending, and this task is accomplished through the use of risk assessment tools. Determination of sex offenders' risk for reoffending is a critical step towards understanding how they are different from the everyday offender. Accurately assessing offenders' potential risk for reoffending can save money, protect citizens and advance knowledge about sex offenders.

According to the Bureau of Justice Statistics (2009:1,5), 1.3 percent of all state felony arrests in 2006 were for sex offenses, with the vast majority of those offenders sent to incarceration. The Bureau of Justice Statistics (BJS) also published a sex-offender recidivism paper in 1994 that outlined recidivism risk and actual post-corrections outcomes for over 9000 sex-offenders released from prisons in the US in that year. The study followed each offender for three years after their initial release. Forty-three percent of the sex offenders were rearrested for a non-sex crime within that three years, and 5.3 percent were rearrested for a sex crime (BJS 1995:7-8).

The reoffending characteristics of the offender group studied by BJS suggest that a re-offense with a general criminal violation is much more common than sexual reoffending, yet the latter is still more troubling to the general public than the former. Despite this general

finding, sex offenders are four times more likely than non-sex offenders to reoffend sexually, and understanding the forces that may cause sex offending is necessary to identify those who possess the highest likelihood for repeating such behavior. Other sources cite overall recidivism rates as low as 13.4%, while individual long-term studies (more than 20 years after an offender's discharge) have discovered that reconviction rates are greater than 40% (Hanson and Bussière 1998:348, Prentky et al. 1997:635).

Findings such as these should not be treated as final truths, given the fact that there is no uniform definition of recidivism across correctional jurisdictions. Some jurisdictions may define recidivism as a return to prison for any reason, whereas others may only be concerned with a return to prison for a sexually-based reoffense; thus any reports of sex offender recidivism may not reflect the actual type of reoffending that has taken place (Hanson and Bussiere 1998:350).

In order to fully address the topic of risk assessment, some preliminary discussion of the supposed causes of sex-offending is necessary to understand where items used in risk appraisal scales originate. Another important step is discussion of what risk appraisal or assessment means. This paper will also provide a detailed inventory of several of the most widely used risk assessment scales, including reviews of recent academic research evaluating each scale. Careful systematic review of the risk assessment scales will highlight the pertinent issues present in risk assessment as a whole, and with individual instruments. The issues present in risk assessment can be overcome, and this review offers solutions to aid the use of risk assessment and the progression of sex offender theory.

CAUSES/CORRELATES OF SEX OFFENDING

Risk assessment instruments are drawn from our knowledge of supposed predictive characteristics, being able to identify these characteristics will help stakeholders understand how scales are constructed. Many of the traits and characteristics discussed in this section have been operationalized as measurement items in risk assessment scales, therefore it is important to discuss what they mean and how they are specifically related to offending.

Kirsch and Becker's 2005 (212) sex offender meta-analysis showed that traits like low self-esteem, previous sexual abuse of the offender themselves, low socioeconomic status, impulsivity, and minority status are common among sex-offenders, yet are more likely to be initiating forces rather than those that maintain behavior. Initiating forces are thought to start beginning patterns of sexual offending, while maintaining forces are thought to keep the deviant thoughts and behaviors of sex-offenders going throughout their lives, after they have been initiated.

Kirsch and Becker's meta-analysis also provides evidence suggesting that the correlates of general criminal behavior are also the most closely associated characteristics with sexually-based offenses. This is problematic for current popular modes of thinking; popular logic leads people to believe that sex-offending is accurately predicted by its own set of distinct characteristics. Kirsch and Becker discovered that variables related to general anti-social attitudes and thoughts were the strongest predictors of both general and sexual recidivism. Such correlates included unstable employment, involvement in crime at an early age,

association with other criminals, alcohol/drug use and pro-criminal belief systems (Kirsch and Becker 2005:212).

Impulsivity is a correlate of general criminality thought to be a strong predictor of sexual offending and sexual re-offending as well. One of the popular beliefs about sex-offenders is that they engage in careful and extensive planning prior to the commission of their acts. As with general offenders, this idea is not well supported by recent meta-analyses. Kirsch and Becker state that like most kinds of crime, many instances of sexual offending may be truly impulsive (Kirsch and Becker 2005:213).

Another trait closely related to sex-offending is lack of empathy. The logic behind measuring this trait is that anyone capable of committing acts of such abusive nature towards other humans must be incapable of having any feelings. This notion is not widely supported, as researchers have found that sex-offenders are indeed able to recognize harm done to others (Marshall, Hamilton and Fernandez 2001:128). According to Marshall, Hamilton and Fernandez (2001), sex-offenders still showed deficits in empathy expression for their own victims, and even to whole “categories” of victims. In other words, offenders were able to show some empathy, but not for victims or those who could be potential victims. The study also showed that sex-offenders (child molesters in particular) showed distorted empathy views on the subject of adult-child sexual relationships (Marshall, Hamilton and Fernandez 2001:129).

One more frequently cited correlate of sex-offending is low self-esteem. Low self-esteem has proven difficult to pin down as a definite predictor, as it is often the outcome of several other predictors of general criminality and sex-offending (e.g., divorce, unemployment

and age are often sound determinants of self-esteem, and are also highly correlated with sex-offending) (Kirsch and Becker 2005:214).

These supposedly well-known correlates of sex-offending have found their way into actuarial risk scales for determining reoffense risk, as will be discussed below. Some characteristics are no longer strong predictors once they are combined with other items in these scales. This may be evidence that interacting relationships between these correlates exist. Interactions in this sense can be thought of as mitigating or aggravating for example. In other words, the presence of some risk factors may mitigate or lessen the effect of other risk factors present offenders. Likewise, some risk factors may amplify or aggravate other risk factors. The following sections on the history of risk assessment and the inventory of available risk scales highlight how some predictive characteristics may look good on paper, but may not sufficiently explain the underlying issues with regard to reoffending.

RISK ASSESSMENT TOOLS

Risk assessment scales form one primary component of sexual offense treatment, and without a determination of a sex offender's risk for reoffense, treatment providers cannot provide the best possible options for patients. Risk assessment was first used in the parole process, to gauge the likelihood of parole success. Consideration of such risk was conceived in the early 1920s by S.B. Warner and Hornell Hart, who believed that information readily available to judges and officials in Massachusetts could be better utilized in order to prevent parole failure among offenders of all types (Hart 1923:405). This early study successfully recognized and confirmed the existence of criminal risk factors that are widely accepted to this

day, such as family disruption, prior parole violation, prison misconduct and mental illness (Hart 1923:407). Hart believed that the total effect of all such predictors should be combined into scores that predicted offender's success at parole. Previous attempts at predicting successful parolees merely inquired about several weak indicators separately, and no scores were calculated (Hart 1923:412). This in itself shows how necessary it is for policy makers and practitioners to use empirical validation for any element incorporated into a risk scale, as relying on weak indicators hinders the development of sound assessment.

Decades later, risk assessment still follows the same ideas that Hart first articulated in the 1920s, with considerable improvements upon his concepts of risk. One of the main improvements to risk assessment strategies is the distinction between two major kinds of risk factors: static and dynamic. Static factors are those risk factors that are relatively fixed within the person; traits like poor adolescent adjustment or unstable family history. Static factors operate like Hart's initial conception of risk factors, as characteristics that are relatively stable and readily observed through several information sources. Dynamic risk factors are theorized to be better suited toward predicting specific types of offending (as well as reoffending) than static risk factors. Dynamic factors are elements capable of changing over time, and can be divided into two subtypes: acute and stable factors (Hanson 1998:3).

Acute risk factors are aspects of life that change rapidly (like sudden onset of drunkenness or sexual arousal) whereas stable risk factors incorporate the same notions over longer periods of time (like the development of deviant sexual beliefs or alcoholism). By distinguishing between stable and acute risk factors, risk assessment can focus on stable factors

to identify long term problem behaviors, or focus on acute risk in order to predict the timing of subsequent offenses (Hanson 1998:2)

R. Karl Hanson (1998:4), of the Canadian Solicitor General's Department, believes that there are three fundamental philosophies for creating risk assessment. First, a multitude of empirically validated risk factors are used to generally describe offenders' reoffense risk in what is called the guided clinical approach. Second, an actuarial approach evaluates offenders on a limited number of predictors, and then a score is assigned based on a weighted rubric (much like Hart). Third, an adjusted actuarial approach allows flexibility in which predictors get analyzed through the use of multiple evaluation tools

INVENTORY OF RISK ASSESSMENT SCALES

VRAG (Violence Risk Appraisal Guide)

Several different scales have been developed in the past few decades that have proven useful in assigning risk to sex offenders, including one scale that was originally developed for use on violent offenders (English 2002:80). The Violence Risk Appraisal Guide (VRAG) developed in 1993, was intended for use on violent offenders in general, but tested as a potential risk tool for sex offenders. The instrument was created during a study of over 300 male inmates at a maximum security psychiatric facility in Canada (Harris, Rice, and Quinsey 1993:317). The inmates' files were examined and coded on several variables, including: childhood maladjustment (scored from 0-none to 3-serious discipline or attendance issues), teen drinking abuse (scored from 0-never drank to 3-serious drinking problem), separation from parents (either yes or no due to divorce abandonment, etc), aggression (scored from 0-none to

7-frequent extreme aggression), and victim injury level (scored from 0-none to 7-death with body mutilation). The study also used a “level of supervision inventory,” a modified version of the Psychopathy Checklist-Revised and criminal history reports (Harris, Rice, and Quinsey 1993:321,323). Despite the instruments’ inclusion of many validated correlates of crime, a very weak relationship between VRAG scores and sexual offense recidivism was observed (English 2002:80).

A 2009 study conducted with Scottish correctional populations tested the predictive validity of the VRAG against other violent recidivism measures. After assessing case files of inmates at a Scottish “clinic” for mentally defective violent offenders, the researchers found that the VRAG was the best overall predictor of violent behavior, particularly serious violence (Ho, Thomson, and Darjee 2009:536). One cited drawback Ho and associates mentioned was the VRAG’s lack of dynamic risk indicators. The authors contend that the VRAG would make a useful initial measure to establish baseline violent tendencies to aid other instruments or treatment (Ho et al. 2009:537). According to the results tables, the VRAG showed the highest mean estimates of risk; something the authors stated as a potential reason why the VRAG might overestimate risk potential (Ho, Thomson, and Darjee 2009:535).

A similar study conducted in England by Doyle, Dolan, and McGovern found that the VRAG showed poor predictability of violent recidivism. Similar to Ho, Thomas and Darjee’s study, the reason stated for its failures was a lack of dynamic indicators and its heavy reliance on items borrowed from the Psychopathy Checklist-Revised (Doyle et al. 2002:151). Note that the PCL-R was designed to measure psychopathic tendencies, not sexual offending likelihood.

No information about VRAG's ability to predict sexual recidivism was collected by Doyle and his colleagues.

As evidence of the VRAG's applicability to sex-offenders, a 2000 Swedish study tested the VRAG in light of the problems mentioned and its purported "overly simplistic nature" (Grann, Belfrage and Tengström 2000:99). Grann and associates' study used a sample of male violent offenders from a Swedish clinical corrections setting and were tested on the VRAG. Grann, Belfrage and Tengström then collected recidivism measures two years after release from the institution (Grann et al. 2000:100). After separating the offender sample into two groups, those with personality disorder and those with schizophrenia, the VRAG was tested in the samples, along with recidivism rates after 2 years (Grann et al. 2000:103). The study found that even though some items in the VRAG could not be completed (notably the time consuming PCL-R items), it possessed moderate predictive validity in their sample of mentally disordered offenders (Grann et al. 2000:110).

Given the limitations of the VRAG in predicting a wide array of recidivism type, it is apparent that an evolved version of this scale could be affective if it were targeted toward sex offenders, by including some sex-offender specific items. Building on previous scales aids the process of creating new instruments, especially if previous scales show some promise with revision. The VRAG is a moderate predictor of violent behavior that could benefit from adding sex offender-specific items.

SORAG (Sex Offender Risk Appraisal Guide)

The Sex Offender Risk Appraisal Guide (SORAG) utilizes ten items from the VRAG but incorporates four other sex-offense specific items, including a phallometric measurement item that covers deviant sexual interests (English 2002:80). Scores on the SORAG range from -27 to 51, with an expected average of zero (Ducrol and Pham 2006:19). The inmate scores are divided into nine “bins” or categories that correlate with level of risk (Looman 2006:196). The SORAG was designed from the start to predict violent sexual reoffending but it was also thought to be more valid than VRAG at predicting non-violent sexual recidivism, and better still at predicting general recidivism.

A 2006 study conducted by Durcol and Pham retrospectively analyzed data from sex offenders in a maximum security prison during a twenty year period. The SORAG predicted violent recidivism better than it predicted sexual recidivism, and the instrument possessed moderate predictive ability in general compared to other common appraisals (Ducrol and Pham 2006:23). Their study also found that SORAG was better at predicting recidivism for rapists than child molesters. In their concluding remarks, Ducrol and Pham note that the SORAG appears to have greater face validity than some other scales, since the SORAG covers a broad range of measures applicable to sex offenders. The researchers contend that in order for a risk assessment process to be complete, multiple assessment tools ought to be used in conjunction with one another. In this way, the SORAG’s failure to predict non-sexual recidivism can be mitigated with an instrument that is specialized toward non-sexual recidivism (Ducrol and Pham 2006:25).

A 2003 study of Arizona sex offenders compared the SORAG with other common appraisal scales, notably the Static-99 and RRASOR. The researchers grouped study subjects by their offense type (rapists, extrafamilial, noncontact and incest offenders) and assessed them on the SORAG, using substitute measures in place of the PCL-R items on the SORAG (Bartosh, Garby, Lewis, and Gray 2003:427). They found the SORAG to have significant predictive ability for all types of recidivism, including general recidivism (Bartosh et al. 2003:430). Bartosh and colleagues concluded that the SORAG was the least predictive of the scales tested when it came to sexual reoffending, but was still a moderate predictor of other types of recidivism. With regard to offender type, the SORAG was most effective for use on rapists and least effective with non-contact offenders (for example, individuals who commit internet-based soliciting types of crimes) (Bartosh et al. 2003:434). The authors admitted that their SORAG results may not have been consistent with previous research because they had to omit items from the SORAG that could not be obtained from prison records, such as child attachment and phallometric measures (Bartosh et al. 2003:435).

The study of Bartosh and colleagues brings up one key issue of contention that other researchers have had with using the SORAG: it is time consuming to acquire the content needed complete the SORAG (Kingston, Yates, Firestone, Babchishin, and Bradford 2008:469). In Bartosh and colleagues' study, they simply omitted four items in the SORAG and put zeros in place of the actual scores. The SORAG only has fourteen items; deleting any of the items reduces its predictive validity. As with other scales that use the PCL-R, the PCL-R items seem to also contribute to difficulties in completing the scales. Kingston and coauthor's 2008 study also chose not to include the childhood maladjustment item out of convenience. The study affirmed

the previous study's findings that the SORAG predicted general and violent recidivism better than sexual recidivism (Kingston et al. 2008:475).

Jan Looman's 2006 study also attempted to differentiate predictive validity between the SORAG and the Static-99 scale, using a sample of over 300 high-risk violent sexual offenders (195). Looman's preliminary results showed that a majority of the offenders in the sample fell in the higher end of the nine bins (Looman 2006:199). Looman (2006:202) found that the SORAG scores were uniformly higher among the offenders, and may have overestimated the likelihood of recidivism Overall, Looman (2006:204) concluded that the SORAG adequately predicted a wide array of reoffending.

Barbaree, Seto, Langton, and Peacock's 2001 study of several risk assessment scales confirmed what previous studies have found, and used an archival method as well. The authors in this particular study were able to fully complete all fourteen items in the SORAG for 55 percent of their cases, which increased the validity of the instrument greatly when compared to other studies that simply omitted. (Barbaree et al. 2001:502). The research team, however, was able to gather complete information for 96 percent of subjects when they used the shorter RRASOR instrument (Barbaree et al. 2001:503).

In comparing SORAG scores with the other instruments tested by Barbaree and colleagues, the SORAG was the best overall predictor of general, sexual, and violent recidivism. (Barbaree et al 2001:508). Barbaree and associates also note that in their opinion, the components of the SORAG that are the most time consuming and intensive to complete are the PCL-R items and the phallometric measures. Given that studies measuring recidivism are

usually more retrospective in nature (analyzing records of offenders who have already recidivated and returned to prison), difficulties with completing the scale are understandable. The PCL-R and phallometric tests are not commonly administered to inmates entering prison, or even most sexual offenders entering residential treatment programs (Barbaree et al. 2001:516).

Like the VRAG, the SORAG includes items that are time consuming and potentially difficult to complete fully. While the instrument has shown effectiveness in predicting sexually-based reoffenses, it still is not generalizable to all offenders and all kinds of recidivism. The problem of being able to generalize results to all kinds of recidivism is one that plagues all risk assessments. Creating a simpler scale that is efficient to complete time-wise however, is not so difficult to accomplish.

RRASOR

The Rapid Risk Assessment of Sexual Offense Recidivism (RRASOR) was also built upon previous work in the field. This index is shorter than previous indexes, and according to the creators, shows moderate predictive ability in relation to recidivism (Hanson 1997:7). It does, however, lack an item commonly used in other scales: a measure of deviant sexual interests (Hanson 1997:19). The RRASOR scale was constructed in order to create an easy-to-complete scale that makes efficient use of time. Hanson argued that current risk appraisal practices lacked efficient means for determining risk (Hanson 1997:4). The method of building the RRASOR differed from what was used to develop the SORAG; since the SORAG was simply an

extension of the VRAG. Developing the RRASOR utilized new items from different sources as well.

Eight potential predictor variables were conceptualized after an overview of Hanson and Bussiere's 1996 meta-analysis, retaining items that had at least a 0.10 correlation coefficient with a recidivism variable. The eight items included: prior non-sexual offenses, stranger victims, marital status, and non-related victims, number of prior sex offenses, age, extrafamilial victims, and male victims (Hanson 1997:6). Hanson measured recidivism as any sexual reoffense, and the initial eight items were tested on seven different offender samples that represented a variety of Canadian correctional institutions where sexual offenders may reside (Hanson 1997:6). After running the eight items against recidivism in the different sample institutions, the four items with the highest correlation coefficients were selected (Hanson 1997:13). As such, four items encompass the entirety of the RRASOR: number of prior sexual offenses, age (specifically whether the offender is under age 25), extrafamilial victims, and male victims. Hanson then backed up this method with a regression analysis, which provided statistical verification of the four items. The samples were re-analyzed using the new four item measure, and they showed moderate predictive accuracy in predicting sexual recidivism in the retrospective samples (Hanson 1997:14-16).

Hanson also discussed future implications of his study, some these implications have been addressed by subsequent research, while others have not. Since the RRASOR was designed to measure risk of sexual recidivism, Hanson thought that using a series of similar short scales could be used to provide measures of risk for all categories of recidivism. These

“subscales” could be developed for subtypes of sex offenders as well (such as child molesters vs. non-contact offenders). To accomplish this, different predictors could be assessed according to offender type. Age for example, may be more predictive in child molesters than rapists, and would therefore not be assessed in a scale tailored to rapists (Hanson 1997:18). Hanson also concluded that while the RRASOR is indeed predictive, it fails to include many sex-offender specific variables, such as deviant sexual practices. Nonetheless, this scale is intended to be a brief, efficient indicator of potential recidivism.

Hanson’s concludes that the chief benefit of the RRASOR scale is its ability to be easily and efficiently completed using archival prison records or individual inmate files. Hanson’s disclaimer is that the instrument was not designed to be used in isolation; as with many other researchers in the field of sex-offender assessment, Hanson agrees that combinations of indicator scales may provide the best possible picture of a particular inmate’s post-corrections recidivism risk (Hanson 1997:19). The RRASOR is unfortunately most often tested against other scales as a stand-alone instrument.

Several of the previously discussed studies also evaluated the predictive validity of the RRASOR, including Bartosh and coauthors’ study. This study did take one of Hanson’s suggestions into consideration when evaluating the RRASOR: that future evaluations of the assessment tool should take into consideration varying kinds of sex offenders (Bartosh et al. 2003:424). Using the same sample of Arizona sex offenders, Bartosh and colleagues were easily able to complete the 4 item RRASOR scale, as compared to the SORAG, where components of the scale were missing (Bartosh et al. 2003:427).

According to Bartosh, Garby, Lewis and Gray's analysis, the RRASOR on its own showed the most significant predictive ability for sexual recidivism, yet not as consistently as the other instruments in the study did (Bartosh et al. 2003:429). When looking at the different subtypes of sexual offender, the RRASOR failed at predicting sexual recidivism in extrafamilial molesters, a finding that suggests that significant differences exist in the spectrum of sex offenders. This also indicates that different risk assessment tools ought to be applied to different types of offenders. The RRASOR was also ineffective when it came to incest offenders, and failed at adequately predicting sexual recidivism in non-contact offenders or rapists (Bartosh et al 2003:430). Bartosh and colleagues concluded that the RRASOR was not valid as a predictive tool for violent offenders, but that it may be true to-its-word as a time-saving preliminary assessment tool for correctional use (Bartosh et al. 2003:433).

Langton, Barbaree, Seto, Peacock, Harkins, and Hansen (2007:37) also tested several risk appraisal instruments, among them the RRASOR. While the main purpose of their 2007 article was to test the predictive accuracy of the Static 2002, their work includes useful conclusions about the RRASOR. Like many of the articles previously discussed, Langton and colleagues' goal was to independently validate the RRASOR and compare it to the multitude of instruments used in risk assessment.

Based on Hanson's initial conclusions, Langton and his coauthors expected the RRASOR to have the greatest predictive ability among the scales tested (Langton et al. 2007:41). According to their work, the RRAOSR appears to be a surprisingly effective tool for only having four items. Generally, other effective scales are much longer. Data and information for the

study was collected retrospectively from a few different Canadian sources, including the Royal Mounted Police. Four types of recidivism were measured: sexual recidivism, violent recidivism, non-contact recidivism, or any other general re-offense. The average time between release and reoffense in the sample was 5.9 years (Langton et al. 2007:47). In the analysis, the team found that the RRAOSR was not a reliable predictor for general or violent recidivism, yet was one of the most valid predictors of sexual recidivism. The authors also state simply that the RRASOR failed at predicting non-violent recidivism (Langton et al. 2007:49-50). Similar to previous research on the RRASOR, Langton found its sole predictive ability seems to lie in assessing risk for sexual recidivism (Langton et al. 2007:54).

The latest test concerning the RRASOR comes again in the context of a comparison with the Static 2002, this time from a sample of offenders in the American Midwest. The initial sample included about 20 female sex-offenders. The females were excluded from the analysis, but no specific argument for this was offered. Most likely it was due to the fact that the instruments discussed here were developed and initially tested on entirely male samples; significant differences exist in the criminological realm with regard to gender, yet they have yet to be determined or analyzed in this context. Offenders in this study were categorized as violent, nonviolent, and as “domestic batterer” according to criminal records and probation/treatment files of subjects (Stalans, Hacker and Talbot 2010:617-619).

Upon scoring the subjects on the RRASOR, the researchers made one important criticism of the instrument: that the instrument fails to consider the effect of any prior undetected crimes about which probation officers or treatment agents may still have knowledge. The

overall sexual recidivism rate for the sample was essentially equal to the rates found in recent meta-analyses (Stalans et al. 2010:622). Of the scales tested, the RRASOR was not the most predictively accurate for any type of recidivism, and did not make a good showing in light of the offender categorization that the authors used. In other words, the RRASOR may have been shown in earlier studies to be an excellent predictor of sexual recidivism among a wide array of offenders, but when Stalans categorized the offenders, the validity of the RRASOR dropped considerably (Stalans et al. 2010:626). An important feature of Stalans' work was the focus on domestic batterer offenders, a subtype of offender that is difficult to criminally prosecute and whose true rates of offending are likely undetected.

When looking back on the RRASOR and the research on it, several points can be made. First, the RRASOR is an efficient and quick way to assess potential recidivism risk in sex offender samples, when the samples are not differentiated into subtypes. In the discussed studies, the RRASOR failed to adequately predict general recidivism or violent recidivism, however, the scale was designed from the outset to predict sexual recidivism only. Due to its replicated accuracy in predicting sexual recidivism, much of the RRASOR was combined with elements of another scale, the SACJ-Min (Structured Anchored Clinical Judgement-Minimum) to create the Static-99, easily the most widely tested and used risk inventory.

What research on the RRASOR shows is that there needs to be a careful balance between creating short, timely instruments and those instruments that are much more involved (and thereby less efficient). The SACJ/SACJ-Min was developed shortly after the RRASOR, and took advantage of a step-wise approach. A step-wise approach to risk assessment

allows for adjustment in offenders' risk scores as they serve sentences or progress through treatment.

SACJ/SACJ-Min

The SACJ was revolutionary with its multistep procedure, but after it was combined with the RRASOR to make the Static-99, it was not commonly used. The SACJ is a step-wise assessment tool, one that initially gives broad risk assignments to offenders and then further specifies the offender's potential for future harm. Information about Thornton's original paper on the development of the SACJ was difficult to uncover, yet Grubin's 1998 review of Thornton's work sheds some light on the development of the measure. The instrument originated in the British Isles under Her Majesty's Prison Service as part of England's national sex offender program (Grubin 1998:36). Grubin argues that the SACJ is not solely dependent upon archival data like other instruments and can be modified and added to over the course of an offender's lifetime in a correctional system.

The SACJ assessment process consists of three steps that occur during an offender's course of treatment. The first step is a broad determinant of the offenders risk, scored as either high, medium, or low risk based on three key characteristics: current and past sex offenses, past and current violent offenses, and more than three convictions of any other sort (Grubin 1998:37). The second step in the process takes aggravating and mitigating factors into account, which would increase or decrease the offender's score from step one. The second step in assessing offenders includes: score of 25+ on the PCL-R, never married, deviant sexual arousal, male victims, stranger victims, substance abuse, ever been in care, and non-contact sex

offenses. The third step in the appraisal process occurs during an offender's progression in corrections, where an offender's risk is increased if they display non-compliant behavior in treatment, and their risk level is decreased if the offender displays acceptable progress in a sex offender treatment setting (Grubin 1998:38). In this way, the SACJ measures dynamic factors.

The SACJ can be completed and evaluated without completing the third step; in this iteration it is known as the SACJ-Minimum. Thus, the SACJ scale is not limited by being a single attempt at determining an offender's risk; scores have the ability to change over time, reflecting an offender's progress or deterioration in treatment (Grubin 1998:38). Grubin also addresses a point reiterated by many other researchers in the field: that while the instruments themselves can be valid predictors, they ought to be used in conjunction with one another for the most accurate and complete picture of one's reoffense risk (Grubin 1998:39).

Initial testing of the SACJ was conducted on a longitudinal sample of 533 British inmates, re-measured 16 years after their release. The SACJ was able to distinguish three groupings of sex-offenders based upon their level of risk: a low risk group in the sample that reoffended at a rate of 9%, a medium risk group that reoffended at a rate of 23%, and a high risk category that reoffended at a rate of 46% (Grubin 1998:38). Grubin stated in his paper that three things needed to be addressed at that point in time in relation to risk prediction. Most importantly, that frequency of offending versus severity of future offending was not distinguishable by (then) current instruments. He also stated that differentiating the samples into three risk categories was a good idea, but differentiating the middle category further would be ideal, and

finally that the instruments may not be applicable for identifying first-time sex-offenders, but is useful for predicting their future offending (Grubin 1998:39).

Systematic test results of the SACJ are difficult to obtain, yet a 2004 study of UK sex offenders sought to compare its predictive abilities against other popular scales (including the VRAG and Static-99). The study was conducted on a small sample of 172 inmates in a medium-risk psychiatric facility, and the researchers were only able to complete the first two steps in the scale. They did not complete the third portion, where inmates' progress is tracked over their treatment/prison career (Craig, Browne, and Stringer 2004:11). The purpose of their analysis was to compare how the same group of sex-offenders was rated differently by each of several instruments they tested. Given the similarity in included items between the different scales (including the ones discussed thus far in this paper), it was not surprising to read that scores on different scales all correlated highly with one another (Craig et al 2004:16).

Compared to the other instruments, the SACJ-Min showed the greatest variability because of its apparent tendency to overestimate recidivism potential in the high-risk category (Craig et al. 2004:19). Few specific conclusions about the SACJ-Min were made in the article, save for some discussion of its limitations; because the SACJ-Min makes use of the PCL-R, it is more difficult and time consuming to complete than other instruments. In the case of Craig and his associates' study, they elected to pick out psychopathic traits from inmates' files in order to come up with a score akin to the PCL-R (Craig et al. 2004:21). Some difficulty expressed in the Craig article may have been due to their small sample size as well.

The authors conclude that future directions in the field of risk assessment would benefit from incorporating more than just static risk indicators (although given the opportunity to complete the dynamic items in the SACJ, Craig and colleagues declined when evaluating the scale). Craig and his associates make many of the same arguments that scale developers have made with regard to offender type as well. They believe that successful instruments should be able to not only distinguish between varying offender types, but also distinguish the different types of recidivism.

After the initial completion of the Static-99 instrument, Thornton and Hanson decided to test the predictive validity of the tool against the scales that compose it (as briefly mentioned earlier, the Static99 is basically a combination of the SACJ-Min and the RRASOR) (Hanson and Thornton 2000:121). Even though the SACJ has the capability to measure dynamic risk factors in its third step, they elected to test the SACJ-Min. The samples that Hanson and Thornton used were the same samples that Thornton used in his initial testing of the RRASOR (Hanson and Thornton 2000:124).

As the authors expected, the combination of the RRASOR and SACJ-Min into the Static-99 produced more accurate results than either the RRASOR or SACJ-Min did alone. Thornton and Hanson concluded that each separate scale is valid for different characteristics; thus, combining the two scales into one would avoid the problem of overlaps (Hanson and Thornton 2000:129).

Static-99

As mentioned in the section under SACJ, the Static-99 was created by combining ten items from both the RRASOR and the SACJ-Min to create a static risk assessment tool that was equally applicable to all types of sex offenders and recidivism categories . To review, the RRASOR included the following items: prior sexual offenses, age (specifically whether the offender is under age 25), extrafamilial victims, and male victims (Hanson 1999:8). According to its developers, the Static-99 has more predictive validity than either the RRASOR or SACJ alone (Hanson and Thornton 2000:129). Dozens of studies around the world have been conducted on the predictive validity of the Static-99; this section will not be addressing all of these, but rather review the newest studies to find a saturation of common issues and problems with this scaled instrument.

The RRASOR, SACJ and the newly formed Static-99 were tested against each other using three archival sex-offender samples from England and Canada (Hanson 1999:6). In all categories of recidivism, the Static-99 predicted recidivism better than the other instruments on their own (Hanson 1999:11). The Static-99 groups offenders into four risk bins: low, medium-low, medium-high, and high. This procedure addresses Grubin's desire to have the middlemost risk category divided up further. In Hanson's discussion, he states that the Static-99 predicts both general and sexual recidivism in a way that neither the SACJ-Min nor RRASOR can. Hanson also believed that a future version of the Static-99 could incorporate more static risk factors (specifically those that measure deviance or persistence of offending), or even include dynamic risk factors like the full version of the SACJ does (Hanson 1999:15-17).

Barbaree and colleagues' 2007 cross-validation study pitted the Static-99 against several other scales, among them the RRASOR, VRAG, and the SORAG. The study also included testing the PCL-R as its own predictive measure (recall that the PCL-R is a component of the VRAG, SORAG and the SACJ). Even though the PCL-R was not developed for use in predicting recidivism, the authors state that the tool is useful for such purposes (Barbaree et al. 2007:495-496).

In Barbaree and colleagues' sample, the average age of offender was 37 and average socio-economic status much lower than average. While the authors distinguished various types of recidivism when measuring post-corrections outcomes of subjects, they did not take into account the various types of sex-offender. Of the instruments tested by the researchers, the Static-99 was the most "completeable"; in other words, this assessment tool was able to be fully completed for a greater proportion of subjects than any other tool tested in the study (Barbaree et al. 2007:503). The Static-99's ease of use is partly due to the fact that the Static-99 does not include the PCL-R component of the SACJ, which many authors have stated is time-consuming and resource intensive to complete (Barbaree et al. 2001:515).

Scores on the Static-99 were highly correlated with scores on the RRASOR (given that the RRASOR is a component of the Static-99) and the predictive validity statistics of the Static-99 on three kinds of recidivism (sexual, violent and general) were nearly all equal (Barbaree et al. 2001:507). As a whole, the Static-99 displayed high predictive abilities, yet in the area of sexual recidivism it was bested by the RRASOR (which studies have shown is valid only for the prediction of such recidivism). In Barbaree and coauthors' work, the Static-99 appeared to be

an excellent all-around tool for risk assessment in sex offender groups (Barbaree et al. 2001: 513). Given the validity of using the PCL-R in risk assessment, the fact that an instrument like the Static-99 that does not use PCL-R items gives credence to the validity of the Static-99

Ducrol and Pham (2006:20) compared the validity of the Static-99 against the SORAG, and were able to complete the PCL-R items (which require structured clinical interviewing techniques). Ducrol and Pham categorized offenders into one of two categories, child molesters and rapists. Their research showed that the Static-99 was slightly more predictive for child molesters than for rapists (Ducrol and Pham 2006:23). Ducrol and Pham suggest that while the Static-99 may be more efficient and less time-consuming for researchers/evaluators to use, it comes across as simplistic when compared to assessment tools that use clinical psychological measures such as the PCL-R. Once again, they mention that risk assessments would benefit with the inclusion of dynamic risk factors, and multi-step application (Ducrol and Pham 2006:25).

Bartosh and colleagues' study of Arizona sex offenders also incorporated the Static-99 into its cross-validation analysis, in addition to most of the previously discussed instruments. Like the other studies of risk assessment scales conducted in the 2000s, the researchers elected to differentiate the offenders in the sample by their offense type (as highlighted in the SORAG section of this professional paper). Recidivism was also broken down into categories, including sexual, general, and serious (violent) reoffending (Bartosh et al. 2003:427).

With regard to the offender sample as a whole, the Static-99 was a significant predictor of all types of recidivism (Bartosh et al. 2003:433). Like most other articles reviewed, none of

the scales tested showed significant predictability for non-contact offenders (Bartosh et al. 2003:433). In conclusion, Bartosh and associate's study recommended the use of the Static-99 in adult and child-victim offender groups because it excelled at predicting violent, sexual and general recidivism (Bartosh et al. 2003:436).

Looman's 2006 analysis also compared the Static-99 with the SORAG on a sample of treated high-risk sex offenders. Looman believed that the evaluation literature at the time was lacking sufficient testing of the Static-99 and the SORAG on a offenders that were already known to be of high risk (Looman 2006:192). When assessed on the Static-99, Looman's offender sample averaged in the "Medium-High" rank category, yet over 40 percent of the sample scored in the "High" category (score of 6 or above).

Looman found that the Static-99 was a valid predictive instrument for offender groups that are known to be high-risk offenders, yet mostly for predicting sexual recidivism (Looman 2006:203-205). It is interesting to note that in several studies, the SORAG predicted a wider array of recidivism than did the Static-99. The primary difference between these two scales is the SORAG's inclusion of the PCL-R. Offenders known to be at the high end of risk for recidivism may have a greater tendency toward psychopathic/sociopathic tendencies; therefore, the SORAG may have more predictive validity due to using the PCL-R in its measurement.

As already discussed, Kingston and colleagues' 2008 study compared the Static-99 with the Risk Matrix 2000 (RM2000), omitting items from the Static-99 that could not be completed with their retrospective data set. Although Kingston's research was heavily focused on

assessing the relatively uncommon RM 2000, the research supported previous findings that the Static-99 was a significant predictor of sexual recidivism, and an adequate predictor of general or violent recidivism (Kingston et al. 2008:479).

Since the Static-99 has been subject to more testing and research than other assessment scales on the market, it has also received more backlash from where these scales fall short. Geneviève Parent, Jean-Pierre Guay and Raymond A. Knight (2011) offered many critiques of how scales are tested; first they believed that the concept of recidivism was more varied than any previous researchers thought. The authors also believed that the use of a single current sample was more beneficial than several retrospective samples, and that as many scales as possible should be tested simultaneously to see where relationships lie between different instruments (Parent et al. 2011:189).

Parent and colleagues also believed that sex offenders could be categorized and differentiated into more subgroups than has previous research. They argue that only one study (Bartosh et al. 2003) has managed to adequately break the offenders into detailed subtypes (Parent et al. 2011:190). Their other criticisms of previous research focus in the follow-up measurements of recidivism, and claimed that longer follow-up periods with evenly spaced measurement intervals (every 5 years for 25 years, for example) are needed to make accurate assessments. This may help because dynamic risk indicators have the potential to change over the course of an offender's lifetime. Using several analytical tools, Parent and colleagues discovered that the Static-99 was not significantly predictive of violent recidivism of rapists, but was significantly predictive of sexual recidivism of child molesters and rapists (Parent et al.

2011:198). Interestingly enough, the PCL-R was tested in their study as well, and found to be a statistically significant predictor of all kinds of recidivism in child molesters. No results were reported about mixed-victim offenders. The authors contend that while some scales focus heavily on predicting either violent or sexual recidivism, the Static-99 offers an effective balance of each type of risk characteristic (Parent et al. 2011:204).

Static-2002

The Static-99 may be one of the most widely used and evaluated risk appraisal tools, yet based on Parent's and others' reviews, it is not without shortcomings that require improvement. Based on the critiques already discussed in the section above, Hanson and Thornton (2003) revised and edited the Static-99 into the new and improved Static-2002. One of the primary issues driving the revision was that the Static-99 was simply the result of combining two smaller, but separate scales. This resulted in what Hanson and Thornton (2003) describe as a somewhat disjointed scale because of the change in item style and measurement as the scale is read. Inter-rater reliability could be improved if the scoring criteria for items in the scale were uniform throughout the entirety of the new version as well. Also, in rare circumstances, it was uncovered that it was possible to have counterintuitive scoring results for offenders. Hanson and Thornton's example of this was how the score for a non-sexual violent offender could decrease if they had a subsequent non-violent sexual offense. Even if the new Static-2002 was not substantially more predictive than its predecessor, at least the authors thought that if it was as valid as before, then the revision was successful (Hanson and Thornton 2003:1).

By reconstructing the Static-99, it became more user-friendly and eliminated erroneous items. For example the Static-99 item “has offender ever lived with a lover for two years” may be a statistically related item to later reoffending, yet it is difficult to ascertain from file records and lacks a theoretical foundation. Twenty-two items were selected and organized into five categories: age at release, persistence of sex-offending, deviant sexual interests, range of available victims, and general criminality (Hanson and Thornton 2003:3).

Items from the original list that were not related to sexual recidivism measures on at least a univariate level were automatically discarded, and those that were related were organized into subscales for multivariate analyses. This resulted in a series of preliminary scales that were tested against various offender subtypes, until a scale was developed that was the most universal and applicable to the widest array of offender type (Hanson and Thornton 2003:5). In the initial testing of their final version, the Static-2002 showed similar predictive abilities as the Static-99 in relation to sexual recidivism. The Static-2002 showed some marked advantages over the Static-99 as well; it significantly predicted violent recidivism as well as sexual recidivism.

In accordance with most other studies discussed, Langton and his team in 2007 conducted a retrospective analysis of the Static-2002 on a treatment center sample. The study, compared the Static-2002 against several other common risk scales, including the Static-99 and the PCL-R (Langton et al. 2007:621). Recidivism was coded into four different categories of offense: serious (violent sexual), sexual recidivism, general, and any recidivism that did not fit into the aforementioned categories. Between the different types of recidivism, about 30-35%

of offenders reoffended in some form after a period of four to five years (Langton et al. 2007:622).

The vast majority of the sample was correctly and completely scored on all items of the Static-2002, and the sample was near normally distributed across the scores. The study mentioned the alpha reliability score for the Static-2002, which was reported at .68. Since the items in the instrument are designed to measure separate indicators related to reoffending, researchers would not expect the items to be well aligned. An alpha score reflects how well the items in the scale group together (in this case, items that don't group well together would be the ideal). If the items in the Static-2002 are not highly inter-correlated, then the Static-2002 can be thought of as have wide applicability. An alpha score of .68 shows moderate levels of correlation among the items in the Static-2002. An exploratory factor analysis was conducted on the Static-2002 to determine how the items align themselves, and if they actually measure what they set out to measure. The five component areas of the Static-2002 were differentiated from each other, while explaining over sixty percent of the variation in the sample scores (Langton et al. 2007:624).

In terms of its predictive abilities, the Static-2002 performed better than the other instruments tested, across all four types of recidivism. Taking the analysis further, Langton and his colleagues (2007) divided the offender sample into three levels of potential risk, according to scores on the Static- 2002. Using a survival forecast analysis, they concluded that offenders in the highest risk category (score on the Static-2002 of seven or more) experienced the sharpest decline in numbers from the time of release, meaning that they returned to prison the

fastest. Calculations for the other two risk categories showed different survival rates, evidence that the Static-2002 possesses the ability to distinguish offenders in terms of their potential risk (Langton et al. 2007:630-633).

Similar to the original development study, Langton and colleagues concluded that it was important to measure the internal consistency of the Static-2002 because they believed it was designed to measure recidivism indicators in terms of broader underlying problems. They found that the five-area grouping that the Static-2002 incorporates was held up by the factor analysis, and that it was a robust predictor of overall recidivism in the retrospective sample (Langton et al. 2007:635).

Stalans, Hacker and Talbot's (2010) recent validation study also included the Static-2002 among its test instruments. Given that the Static-2002 has only been in use since 2003, it has not had the same level of extensive testing that the Static-99 and its predecessors have had. Langton and colleagues stated in their concluding remarks that the Static-2002 should not be used in professional capacity when they published in 2007, but rather should be further evaluated or used as a research tool only (Langton et al. 2007:636).

As with nearly all studies of risk assessment, Stalans and associates' study used a retrospective secondary sample to measure the predictive accuracy of the Static-2002 and several other instruments. Offenders in the three subtypes did not differ from one another in terms of the nature of their crimes, but their risk scores on the Static-2002 indicated that there were some differences between the categories (Stalans et al. 2010:618-619). According to other parts of their inferential analyses, "violent propensities" was the most significant

predictor for domestic batterers and other violent offenders (Stalans et al. 2010:623). Non-family victimization was significantly predicted by the criminal history item in the Static-2002, and criminal history was also a significant predictor for familial crimes. Some of these specific findings indicate that the Static-2002 demonstrates in its ability to accurately predict potential risk for various types of offenders (Stalans et al. 2010:624). The study makes no mention in its discussion about the uses of the Static-2002 only, as Stalans and colleagues suggested.

Another very recent study of the Static-2002 was conducted in Canada by Looman and Abracen (2010), and compared the scale against several other actuarial risk tools. Using the same retrospective analysis methods as previously discussed, the researchers categorized their offender sample according to the offender's prior victim type (e.g., those with under aged victims were categorized as molesters).

The Static-2002 proved to be a valid prediction tool when compared against other common risk scales, and like the previous work in evaluating the Static-2002, Abracen and Looman found that it is most effective in predicting sexual recidivism in adult-victim offenders (Abracen and Looman 2010:798-802). Further inquiry into the Static-2002 led the researchers to break up the five subscales in the Static-2002, and see which of those subscales was the most predictive of sexual recidivism. The subscales "Persistence of Sexual Offending" and "Age at Release" were the most significant predictors of sexual recidivism for rapists; the subscale "Deviant Sexual Interests" was the only significant predictor for child molesters (this finding highlights notions that child molesters and adult-victim rapists are inherently different on a psychological level) (Abracen and Looman 2010:804). The authors point out that their results

are much different from the Hanson (2003) origin study, where all subscales were determined to be theoretically valid. Nonetheless, they concluded that the Static-2002 was very effective at predicting sexual recidivism in both rapists and child molesters, and was a moderate predictor of other kinds of recidivism (Abracen and Looman 2010:806).

CASE IN POINT: RISK ASSESSMENT IN COLORADO

With the wide array of valid predictive tools available, one would think that further development should produce an update of the Static-2002, given that the scales presented in this paper follow a somewhat logical evolution. The purpose of discussing the state of Colorado's sex offender risk assessment scale is to illustrate a few issues with their approach. First, the methods used to create this scale are not necessarily better or worse than other scales, but the scale's construction appears haphazard. Second, their method of scale development does not fully take advantage of the large body of work that occurred to create the previously mentioned scales, and thirdly that no academic reviews of this scale could be discovered that tested the Colorado scale's predictive abilities (or lack thereof).

The state of Colorado decided to design, create and implement their own version of a risk appraisal tool, using items borrowed from other risk inventories, and data from clinical interviews. This method of creating a scale involved the initial research team making a lengthy list of potential indicators and testing them against a sample of offenders, in order to determine which are most predictive. The final scale contains the ten-most highly related items from the original selection. In their initial analysis, 494 adult male convicted sex offenders in

treatment programs in the late 1990's were used to create this assessment scale (English 2002:81).

Extensive research to test applicable ideas is necessary for creating theoretically parsimonious risk scales, and even then years of revision and evaluation still cannot create tools that are perfect. It is this author's opinion that using older meta-analyses to identify potential items and test them on one sample is essentially ignoring the years of research and development that were invested creating previous scales.

Static risk factors addressed in this scale include demographic information, criminal history and family history variables (English 2002:82). The Colorado sex offender risk assessment scale also includes dynamic characteristics. The dynamic indicators used include: empathy, willingness to change, motivation, denial, social skills, interpersonal competence, deviant sexual preferences, treatment compliance and positive social support. All of the static and dynamic items in the initial survey instrument were collected via the treatment providers who worked with subjects (English 2002:84). As mentioned in the discussions of the scales above, dynamic indicators are thought to hold the key to making the most predictive scales possible. Those scales that utilize these kinds of factors appear to have excellent effectiveness, and strong conceptual validity.

Risk of failing treatment was the study's outcome measure, and included treatment revocation, pending revocation, treatment termination, commission of a sex crime, "absconded" or even being on the brink of failure according to the therapist. The final scale itself as used in the Colorado correctional system consists of ten items, tested in the original

study at 12 and 30 months after treatment (English 2002:85). The ten items in the scale are described as follows:

- **Adolescent Delinquency.** Similar to other risk scales, the Colorado risk assessment scale addresses the offenders' prior criminal history and deviant activities of their youth. Adolescent deviant behavior and formal juvenile delinquency is accounted for in the scale because of the well established criminological connections between criminal involvement at young ages with adult criminality.
- **Felony Convictions.** Prior adult felony convictions are addressed for the same reason that juvenile criminality was included: past behavior is a strong predictor of future behavior (English 2002:89).
- **First/Second Grade Failure.** The Colorado scale also asks whether or not offenders failed the first or second grades. The rationale behind the inclusion of this item follows the same logic as item one: early onset of problem behaviors is predictive of conduct disorder, and conduct disorder is related to juvenile delinquency.
- **Employment.** The developers were concerned with whether or not the offender was employed full time preceding their initial arrest. The authors of the Colorado scale cite a few different reasons for this item's inclusion, the most important was that fully-employed persons have less free time to commit offenses, or that the higher level of functioning required to keep employment reduces the likelihood that an individual will engage in risky behaviors (English 2002:90). This latter point highlights the earlier discussion that sex-offenders have impulsive tendencies like most other kinds of criminal.
- **Alcohol/Drugs Present.** Existence of alcohol or drugs during a crime, specifically whether or not the victim was under the influence during the sex crime, is also covered in the Colorado scale. The authors stated that this was a key variable in an offender's mode of operation.
- **Sexual Arousal.** Distinguishing violence and power from sexual interest or arousal is also a component used, by asking whether or not the offender was sexually aroused (experienced erection) during commission of the crime. Offenders who reported not being sexually aroused tended to receive longer prison sentences (English 2002:91).

- **Weapon Possession.** The scale asks whether or not the offender had possession of a weapon during the offense, regardless if it was used or not.
- **Denial.** Scores on a denial measurement are also considered by the State of Colorado in determining risk for reoffense. The authors cite Salter (1988) and Brake (1996) as stating that various stages of denial exist and that denial is one of the strongest predictors of reoffense and/or treatment noncompliance.
- **Deviant Interests.** Like some of the previously discussed scales, this instrument also asks about the offender's deviant sexual interests and arousals. In fact, the scale utilizes the same phallometric arousal measures from the SORAG (English 2002:91).
- **Treatment Motivation.** Lastly, the assessment scale measures the offender's motivation for participation in treatment. This motivation element to the scale was borrowed from the Multifactorial Assessment of Sex Offender Risk of Recidivism (English 2002:92). Also known as the MASORR, this scale was not reviewed in the above sections because of its relative lack of research history supporting its use, and it was not widely discussed in the comparisons with other scales.

The Colorado Department of Corrections' decision to include dynamic risk indicators follows the call sounded by many articles reviewed in this paper. Dynamic risk indicators, as previously discussed are thought to be one of the strongest potential risk indicators out there. The SACJ incorporates this type of indicator into their instrument, yet inclusion of dynamic risk indicators increases the cost and time necessary to complete risk evaluations (thus the creation of the SACJ-Min, which eliminated the most costly component: measuring offenders' progress through treatment). Several items used in the Colorado risk scale were conceptualized from items listed in Hanson and Bussiere's famous meta-analysis in 1998. While the Colorado Department of Corrections did implement this risk tool, no research articles could be uncovered that described neither how well the instrument worked in practice, nor whether or not the instrument was still in use.

Despite what the state of Colorado did, such as addressing the need for more dynamic indicators, it still can be viewed as an unnecessary break from the evolutionary process that produced the SORAG, RRASOR, SACJ and the static series of risk assessment scales. Granted, the Colorado risk assessment scale showed moderate predictive abilities for the different kinds of treatment failure, which brings up a key element about Colorado's development study: they did not measure recidivism, but rather treatment non-compliance.

The Colorado risk assessment scale is the only scale discussed which tested for how well treatment non-compliance could be predicted. Risk assessment is a useful step in both the pre- and post-treatment processes for offenders, and while most studies covered in this review measured recidivism, measuring treatment non-compliance comes with some benefits. If risk assessment is intended to aid efficiency in treatment, then refusing to admit those offenders that are unlikely to complete treatment for whatever reason will save time and money, thereby increasing the amount of spending available for offenders who are likely to complete treatment. On the other hand, measuring treatment noncompliance rates does not necessarily equate to reoffending. As stated above, treatment noncompliance was measured in a variety of ways. Ultimately, the ideal testing and implementation of risk assessment scales would measure both recidivism and treatment noncompliance. The State of Colorado has identified a useful approach to testing assessment scales that could spread to other analyses. In this way, clear links between treatment failure/success and later reoffending may be observed systematically.

CONCLUSIONS AND IMPLICATIONS

Based on the state of sex offender risk assessment, the following arguments will highlight possible resolutions to the predictive shortcomings discussed. This review concludes that:

1.) **Evolutionary Process.** Risk assessment scales can be created using previous work as templates, or without using them. Using previous assessment scales as templates is beneficial because more time and research are needed to revise an existing scale than to develop from such scales from scratch.

2.) **Multi-step Scale Application.** If risk assessment tools are not universally applicable to all offenders and recidivism types, then policy implementers should consider evaluating offenders in a multi-step procedure. Because each risk scale is applicable to a certain niche of offenders, using multiple assessments will rate offenders accurately for each potential area of risk.

3.) **Use of Theoretically Grounded Scale Items.** Measures used in risk assessment scales need theoretical grounding in order to help explain causal links that result in sex offending. Risk assessment procedures can test and evaluate theoretical ideas about sex offenders while serving practical corrections needs. Risk assessment practice should also contribute to the advancement of sex-offender and risk assessment theory.

4.) **Using an Evidence-Based Practice Approach.** Scales should be systematically evaluated and organized into listings of evidence-based assessments that would aid

administrators with selecting the most appropriate risk assessment procedures to match the characteristics for their particular sex-offender group.

5.) **Considering Scale Length.** Issues regarding scale size can be mitigated by using multi-step procedures as well. For example, short scales that are predictive of one type of recidivism can be combined with complimentary assessments.

Evolutionary Processes

Hanson and Bussiere's 1998 meta-analysis initiated research into creating risk assessment scales for use on sex-offenders. The resulting multitude of scales was built upon each other in an evolutionary format, where the best elements and ideas were preserved and passed on to successive instruments. This process is aided by having the same researchers work throughout this process, including Hanson. Hanson not only published influential meta-analyses, but he involved in the creation of the SORAG, RRASOR and the Static instruments.

The Static-99 is the result of an addition of two scales: the RRASOR and the SACJ-Min. The SORAG is another example of the influences that previous policy can have on future policy, where the VRAG (Violence Risk Appraisal Guide) was revised with phallometric tests and deviant sexual interest measures to create a sex-offender specific version. The VRAG itself was heavily reliant on items borrowed from the PCL-R (which arguably changed the field of sex offender risk assessment; Andrews, Bonta, and Wormith 2006:22).

Using previous work to create new scales is not without potential problems. Granted many of the scales presented in this review are widely used, they all have weaknesses. When

scales are continually built upon previous iterations, many of the same shortcomings persist. For example, the SORAG suffers from the VRAG's inability to predict general recidivism. However, the SORAG's validation in predicting sexual recidivism offers evidence in support of identifying which scales have particular niches, and using a variety of those niche-specific assessments on offenders will provide a complete picture of potential risk.

Solution-Multistep Scale Application

While having specialized risk assessments provides a partial picture of offenders' risk, the best scenario for assessment is to apply several different instruments. Correctional staff can also reapply those assessments periodically as offenders progress through treatment. This approach will not only provide a more complete measure of an offender's total potential risk, but will show how dynamic indicators change over the course of the incarceration/treatment. The SACJ for example, uses its own multistep procedure for determining an offender's potential risk; after initial determinations of general risk are assessed, dynamic indicators continually measured as the offenders progress in treatment.

Since research has already identified the specialty areas for the most popular risk scales, creating a battery of assessments would not be difficult. For example, using the VRAG to measure violent potential risk, the RRASOR to measure sexual recidivism risk, and the SACJ for general recidivism would provide a more accurate assessment of risk than any scale could on its own. Using multiple measures may result in overlapping predictions or contradictory predictions; no studies were found that tested for these potential outcomes however. Then,

reapplying the aforementioned scales over the course of a treatment program would measure offenders' progress.

Use of Theoretically Grounded Scale Items

Proper scientific methods of theoretical inquiry and hypotheses testing are not the most effective means of creating risk tools within the time and finance constraints that most researchers face. So-called "data mining" techniques, where dozens of items are entered into test models to find the most predictive sets of items works, but doing so leaves the apparent parsimony of the resulting scales in question. For example, the RRASOR was constructed with the four most predictive items borrowed from Hanson and Bussiere's 1998 meta-analysis, and contains items that span the breadth of possibility. Because sex offenders can be dangerous when placed back in regular society, using preliminary instruments (like the narrowly focused VRAG) is advisable for practical purposes, as long as more valid and comprehensive scales are under development. The earliest risk assessment scales were not focused on the ability to differentiate risk in different offender groups, as creating some kind of base scale was the most important task early on.

The ultimate goals for having theoretically-motivated scales are to assess risk as accurately as possible, and to further the development of sex offender theory. Theory provides inferences into the processes that drive sex offender behavior. Theoretically-based assessments will allow for practical testing of ideas about sex offenders. The Colorado Risk Appraisal Scale may have some very predictive items contained within it, but how are they relevant to theory or how universal are they? For example, the Colorado Risk Appraisal Scale

asks whether or not the offender in question completed the first or second grade; while the concept behind the item appears to be assessing the early onset of youth problem behaviors, the item's linkage to sex offender theory is not entirely clear.

Items in risk appraisal tools need to highlight clear linkages between criminological theory, sex offender theory, and patterns of sexually-based offending. If this is accomplished, a risk assessment scale can do more than provide a guess at any particular person's risk; the tools can be used to put theoretical ideas into testable practice. Evidence-based sex offender theories can be developed and evaluated while practical needs are being met in corrections. The incorporation of theory into practice allows for theoretical ideas to be tested hundreds of times, on a variety of inmate samples all across the country. In effect, incorporation of theory would be one of the largest multiple-site replications ever attempted. Large scale theoretical testing will aid in achieving the ultimate goal for any type of correctional programming: a "best practices" approach.

Using an Evidence-Based Practice Approach

Best practices or so called "blueprints" practices are becoming present in the field of juvenile violence prevention. The Center for the Study and Prevention of Violence (CSPV) records and keeps track of successful youth delinquency prevention programs, and requires programs that are included in its list of model programming to meet stringent requirements. The organization monitors and studies program effectiveness to provide the juvenile justice realm with a list of systematically evaluated and proven programs that also serve as models for

new program generation. The requirements for inclusion on their list are: strong research design, multiple site replication, measured sustained effects, and cost/benefit analyses (SCPV).

The risk scales discussed in this paper could be catalogued in terms of their relative successes when evaluated. Sex offender assessment (and perhaps all correctional programming) needs access to information that provides correctional agencies with evidence-based information about available risk assessment scales, and how effective those scales are in predicting offenders' likelihood of reoffending. Evidence-based sex offender risk assessment tools would greatly aid an organization desiring to implement proven risk .Further, access to peer-reviewed sex offender research would provide agencies with more evidence supporting the use of a particular scale. A program designed to offer access to information about scales could also state what types of sex offender groups are best assessed by which scales, and which types of recidivism are best predicted by a particular risk appraisal tool. These may be far-reaching goals, but they are necessary in order to make information about risk assessment easier to obtain. What the Center for the Study and Prevention of Violence has done with creating a website listing effective juvenile programming could be used as a model for how sex offender risk assessments could be catalogued.

Considering Scale Length

Considering the length or number of items in these scales, a scale that only contains a handful of items will be easy and cost efficient to complete, but its predictive accuracy could be called into question. Some of the most highly regard psychiatric instruments like the PCL-R contain many more items than a scale like the RRASROR (with four items) and take considerable

time and expertise to complete and evaluate correctly. When risk appraisal tools use few items, their ability to predict more than one type of recidivism declines drastically. Logically, a scale that is composed of less than ten items will not adequately cover the entire realm of possible predictive characteristics. This can be seen with the RRASOR; it is a relatively short scale that is effective in predicting sexual recidivism, and it does not provide the best possible results for all categories of offenders. This is one reason why the Static-99 was created out of two previous instruments; each instrument demonstrated excellent predictability for specific types of offenders and for specific types of recidivism.

This is not to say that longer is simply better. Corrections departments that wish to implement some type of risk appraisal tool should decide how to balance their needs. Some departments may benefit from multiple scales and extensive assessments, while other departments may need fewer scales and less time, depending on their offender population characteristics. In order for problems associated with sex offenders to be solved, attempts must be made to create the most accurate system of risk assessment. As previously mentioned, some of these scales should be used in conjunction with one another to both provide the most accurate scores for the offenders and contribute to the body of knowledge on sex offenders.

Issues with scale length can also be solved by using the aforementioned multistep procedure. Further research may show that creating a totally universal scale for all types of offenders and all kinds of recidivism is impossible. Designing risk appraisal tools that are specified for offender type and recidivism type will save time when used in a multi-step application. As such, it then would not matter if scales are perceived as being too short or too

long, as long as the scales accurately predict recidivism. Only then, can the goals of applying and modifying sex offender theory, creating model assessment tools, and comprehensive application be achieved.

REFERENCES CITED

- Andrews, D.A., James Bonta and Stephen Wormith. 2006. "The Recent Past and Near Future of Risk and/or Need Assessment." *Crime and Delinquency* 52:7-27.
- Barbaree, Howard E., Michael C. Seto, Calvin M. Langton and Edward J. Peacock. 2001. "Evaluating the Predictive Accuracy of Six Risk Assessment Instruments for Adult Sex Offenders." *Criminal Justice and Behavior* 28:490-521.
- Bartosh, Darci L., Tina Garby, Deborah Lewis and Steve Gray. 2003. "Differences in the Predictive Validity of Actuarial Risk Assessments in Relation to Offender Type." *International Journal of Offender Therapy and Comparative Criminology* 47:422-438.
- Craig, Leam A., Kevin D. Browne and Ian Stringer. 2004. "Comparing Sex Offender Risk Assessment Measures on a UK Sample." *International Journal of Offender Therapy and Comparative Criminology* 48:7-27.
- Doyle, Mike, Mairead Dolan and John McGovern. 2002. "The Validity of North American Risk Assessment Tools in Predicting In-Patient Violent Behavior in England." *Legal and Criminal Psychology* 7:141-152.
- Ducrol, C. and T. Pham. 2006. "Evaluation of the SORAG and the Static-99 on Belgian Sex Offenders Committed to a Forensic Facility." *Sexual Abuse: A Journal of Research and Treatment* 18:15-26.
- Elliott, Delbert. "Blueprints for Violence Prevention" *Center for the Study and Prevention of Violence*. Retrieved April 17, 2011. (<http://www.colorado.edu/cspv/blueprints/index.html>).
- English, Kim, Paul Retzlaff and Dennis Kliensasser. 2002. "The Colorado Sex Offender Risk Scale." *Journal of Child Sexual Abuse* 11:77-96.
- Grann, Martin, Henrik Belfrage and Anders Tengström. 2000. "Actuarial Assessment of Risk for Violence: Predictive Validity of the VRAG and the Historical Part of the HCR-20." *Criminal Justice and Behavior* 27:97-114.

- Grubin, Don. 1998. "Sex Offending Against Children: Understanding the Risk." *Police and Reducing Crime Unit: Police Research Series Paper 99:1-56*.
- Hanson, R. Karl and Monique T. Bussiere. 1998. "Predicting Relapse: A Meta-Analysis of Sexual Offender Recidivism Studies." *Journal of Consulting and Clinical Psychology 66: 348-362*.
- Hanson, R Karl. 1997. "The Development of a Brief Actuarial Risk Scale for Sexual Offense Recidivism." Public Works and Government Services Canada.
- Hanson, R Karl and David Thornton. 1999. "Static -99: Improving Actuarial Risk Assessments for Sex Offenders." Public Works and Government Services Canada.
- Hanson, R Karl and David Thornton. 2000. "Improving Risk Assessments for Sex Offenders: A Comparison of Three Actuarial Scales." *Law and Human Behavior: Advances in Assessment and Treatment of Forensic Populations 24:119-136*.
- Hanson, R Karl and David Thornton. 2003. "Notes on the Development of the Static-2002." Public Works and Government Services Canada.
- Harris, Grant T., Marnie E. Rice and Vernon L. Quinsey. 1993. "Violent Recidivism of Mentally Disordered Offenders: The Development of a Statistical Prediction Instrument." *Criminal Justice and Behavior 20:315-335*.
- Hart, Hornell. 1923. "Predicting Parole Success." *Journal of the American Institute of Criminal Law and Criminology 14:405-413*.
- Ho, Hilda, Lindsey Thomson, and Rajan Darjee. 2009. "Violence Risk Assessment: The Use of the PCL-SV, HCR-20, and VRAG to Predict Violence in Mentally Disordered Offenders Discharged from a Medium Secure Unit in Scotland." *Journal of Forensic Psychiatry and Psychology 20:523-541*.
- Kingston, Drew A., Pamela M. Yates, Philip Firestone, Kelly Babchishin and John M. Bradford. 2008. "Long-term Predictability of the RM-2000:A Comparison with the Static-99 and the Sex Offender Risk Appraisal Guide." *Sexual Abuse: A Journal of Research and Treatment 20:466-484*.

- Kirsch, Laura G. and Judith V. Becker. 2005. "Sexual Offending: Theory of Problem, Theory of Change and Implications for Treatment Effectiveness." *Aggression and Violent Behavior* 11:208-224.
- Langan, Patrick A., Erica L. Schmitt and Matthew R. Durose. 2003. "Recidivism of Sex-Offenders Released From Prison in 1994." Washington, DC: Bureau of Justice Statistics.
- Langton, Calvin M., Howard E. Barbaree, Michael C. Seto, Edward J. Peacock, Leigh Harkins and Kevin. T. Hansen. 2007. "Actuarial Assessment of Risk for Reoffense Among Adult Sex Offenders: Evaluating the Predictive Accuracy of the Static-2002 and Five Other Instruments." *Criminal Justice and Behavior* 34:37-59.
- Langton, Calvin M., Howard E. Barbaree, Edward J. Peacock, Leigh Harkins and Kevin. T. Hansen. 2007. "Reliability and Validity of the Static 2002 Among Adult Sex Offenders with Reference to Treatment Status." *Criminal Justice and Behavior* 34:616-640.
- Looman, Jan. 2006. "Comparison of Two Risk Assessment Instruments for Sexual Offenders." *Sexual Abuse: A Journal of Research and Treatment* 18:193-206.
- Looman, Jan and Jeffrey Abracen. 2010. "Comparison of Measures of Risk for Recidivism in Sexual Offenders." *Journal of Interpersonal Violence* 25:791-807.
- Marshall, W.L., Kate Hamilton and Yolanda Fernandez. 2001. "Empathy Deficits and Cognitive Distortions in Sex Offenders." *Sexual Abuse: A Journal of Research and Treatment* 13:123-130.
- Parent, Geneviève, Jean-Pierre Guay and Raymond A. Knight. 2011. "As Assessment of Long Term Risk of Recidivism by Adult Sex Offenders: One Size Doesn't Fit All." *Criminal Justice and Behavior* 38:188-209.
- Prentky, Robert A., Austin F.S. Lee, Raymond A. Knight, and David Cerce. 1997. "Recidivism Rates Among Child Molesters and Rapists: A Methodological Analysis." *Law and Human Behavior* 21:635-659.

Stalans, Loretta J., Robyn Hacker and Mary E. Talbot. 2010. "Comparing Non-Violent, Other-Violent and Domestic Batterer Sex Offenders: Predictive Accuracy of Risk Assessments on Sexual Recidivism." *Criminal Justice and Behavior* 37:613-628.

US Bureau of Justice Statistics 2009. "Correctional Populations in the United States, 2009." Washington, DC: Bureau of Justice Statistics.