

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Evaluating and Improving Risk Assessment Schemes for Sexual Recidivism: A Long-Term Follow-Up of Convicted Sexual Offenders

Author(s): Raymond A. Knight, Ph.D. ; David Thornton, Ph.D.

Document No.: 217618

Date Received: March 2007

Award Number: 2003-WG-BX-1002

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

<p>Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.</p>

Running Head: EVALUATING RISK ASSESSMENT SCHEMES

Evaluating and Improving Risk Assessment Schemes for Sexual Recidivism:

A Long-Term Follow-up of Convicted Sexual Offenders

Raymond A. Knight, Ph.D.

Department of Psychology

Brandeis University

Waltham, Massachusetts

David Thornton, Ph.D.

Sand Ridge Secure Treatment Center

Mauston, Wisconsin

Evaluating and Improving Risk Assessment Schemes for Sexual Recidivism:
A Long Term Follow-up of Convicted Sexual Offenders

EXECUTIVE SUMMARY

Risk assessment plays a central role in the management of sexual offenders in the criminal justice system, impacting every level of decision-making. The specification of which offenders should be included in state registration or community notification, the distribution of treatment opportunities in prison, the determination of parole eligibility, the allocation of resources in monitoring and supervising offenders in the community, and the ultimate decision to remove the most serious offenders from the community through civil commitment all involve critical choices involving risk. The serious consequences to potential victims, to those offenders who no longer pose a threat to the community, but are denied their civil liberties, and to critical community funds that may be squandered as the result of inaccurate decisions all demand the guidance of the most accurate actuarial algorithms available. Because decisions must be made—even doing nothing constitutes a decision—to choose not to employ the most accurate decision-making algorithms is to choose to pay the costly price of a suboptimal decision.

The aim of the present study was to evaluate and to improve the decision-making algorithms that have been generated to assess risk in sexual offenders. More specifically, it was the task of this research project to evaluate the extant actuarials in a sample of sexual offenders on whom long-term follow up were available. We assessed the comparative accuracy of the major risk instruments over time and over subsamples, explored their underlying factor structure, examined the accuracy of a new assessment protocol (SRA Need Assessment; Thornton, 2002), and explored the potential for generating improved predictive instruments.

In a prior study we (Knight, 1999; Prentky, Knight, & Lee, 1997) had followed 599 offenders who had been referred to the Massachusetts Treatment Center (MTC) for evaluation between 1959 and 1984. Of these 266 (Bridgewater Treatment [BT] sample) had been committed to MTC as “sexually dangerous” and subsequently released, and 333 (Bridgewater Observation [BO] sample) had been determined not to be sexually dangerous and returned to finish their sentences. Of the 333 BO offenders 200 constituted a matched sample (on age at evaluation, marital status, and number of prior crimes) and 200 were randomly sampled from the entire sample BO population evaluated. There was an overlap of 67 offenders selected by both the random and matched process. For all these offenders we had accessed and integrated four outcome record sources, including the Massachusetts Board of Probation records, the Massachusetts Parole Board records, the Massachusetts Treatment Center Authorized Absence Program records, and the Federal Bureau of Investigation (FBI) records.

In the present study we accessed the archival clinical files for these offenders and coded these records both on modern empirically-derived, mechanical actuarials that have been developed since 1997 for predicting sexual recidivism, including the RRASOR (Hanson, 1997), the Static-99 (Hanson & Thornton, 2000), the Static-2002 (Hanson & Thornton, 2003), the SORAG (Quinsey, Harris, Rice, & Cormier, 1998), the MnSOST-R (Epperson, Kaul, Huot, Hesselton, Alexander, & Goldman, 1998), and the Risk Matrix 2000 (Thornton et al., 2003), on two structured clinical guidelines, the SVR-20 (Boer, Hart, Kropp, & Webster, 1997) and the A-SOAP-II (Prentky & Righthand, 2003), and on a new experimental measure, the SRA Needs Assessment (Thornton, 2002). Offenders were also categorized in the Massachusetts Treatment Center Rapist (MTC:R3) and Child Molester (MTC:CM3) typologies and diagnosed according to the DSM-IV Antisocial Personality Disorder and Conduct Disorder criteria. All codings and

judgments were carried out without any access to or knowledge about the follow-up status of the offenders.

The final report addressed six basic questions. We summarize the findings for each using this framework.

1. Reliability and Predictive Potency of Actuarials Applied to the MTC Sample

Although the record sources accessed in the present project were written at a time that preceded the recent research attention to prediction for sexual offenders and consequently did not focus on and detail those variables that have now achieved greater salience among practitioners working in the area, and although these records sometimes provided vague or incomplete prior criminal histories, which are so critical to rating many actuarials, nonetheless the interrater reliabilities achieved by the study's raters and the overall level of predictive accuracy obtained were comparable to other independent evaluations of actuarials (e.g., Langton, 2003). This accuracy was also remarkable in light of the difficulty collecting accurate and comprehensive recidivism data in the USA as compared to the UK, Canada, or Sweden where there are high quality national conviction databases available to researchers. This follow-up problem was likely addressed to some degree because of the strategy in the present study of accessing and integrating multiple outcome sources. Finally, because all of the offenders in the sample had been referred for a legal determination of whether they should be committed to MTC, offenders who were obviously at low risk were likely screened out, and their advantage to enhancing prediction was lost. Consequently, the average risk level of the population was higher than is normally found in a prison sex offender population.

Overall, modern instruments designed for use with sexual offenders showed a moderate level of accuracy predicting sexual, victim-involved, and victimless charges at various time

gates. The large number of predictive instruments and the relatively long outcome period allowed us to examine systematically several issues about the temporal stability of predictors, the differential predictability of subsamples, and the comparative prediction of different crime types. We summarize in turn our results in each of these domains.

Cross-temporal Stability of Prediction

Over the entire sample the average prediction of serious sexual recidivism appeared relatively temporally stable. In contrast, the accuracy of the prediction of victim-involved offenses increased at each time gate examined, and the accuracy of victimless offenses was stable through 10 years, but increased at the 15-gate.

Comparisons between the BO and BT Samples

The present BO sample (those declared not sexually dangerous and not committed) comprised two subsamples, one matched and one randomly selected subsample of all BOs evaluated at MTC. The BTs had both a more rapid and higher rate of recidivism than both the entire BO and the randomly selected BO samples. Estimated recidivism rates from the survival analyses indicated 32% recidivism for the BTs at 15 years and 12.5% for the randomly selected BOs. For victim-involved offenses BTs had more rapid and higher recidivism rates only when compared to the random BOs, not when compared to the entire BO sample. BTs differed from neither the random BO sample nor the entire sample in their subsequent rate of victimless offending.

There is an ongoing debate about what base rates of reoffense should be used as a touchstone to estimate the long-term viability of civil commitment decisions (Doren & Epperson, 2001; Janus & Meehl, 1997). The combined projected 15-year rate of recidivism (21.5%) is arguably the best compromise rate to be used for such considerations for a sample of men being

considered for commitment. This rate corresponds closer to the rates proposed by Janus and Meehl (1997), and suggests that the rates extrapolated by Doren and Epperson (2001) in their criticism of the former article are not appropriate (see the discussion for further elaboration).

In general, and contrary to what might be expected from certain perspectives, BOs serious sexual charges were on average more predictable than BTs, with the differences in average AUCs reaching significance at two time gates and approaching significance on the third. From a Bayesian perspective a smaller probability recidivism target should be harder to predict than a larger one (Barbaree, 1997; Meehl & Rosen, 1955). Consequently, one might hypothesize that the higher recidivism rate of the BTs would allow greater accuracy of prediction. This was not the case. Interestingly, although across the board the means for the actuarials were significantly higher for the BTs than the BOs, the variances of these measures were comparable across the groups, suggesting the possibility that it might be more difficult to predict recidivism among higher risk individuals with the current actuarials. The evidence for slightly higher predictability for the entire sample than for either subsample argues as well that greater heterogeneity of risk might enhance predictive accuracy.

For the non-sexual crimes the BOs tended also to show higher predictability when differences emerged, but the results were not as consistent as they were for sexual crimes. The general conclusion across crime types was that the entire sample yielded the highest AUCs, followed closely by the BOs.

Different temporal patterns in predictability emerged for the BOs and BTs. For all crime types the BOs were fairly consistent across time, showing a slight dip in predictability between the 3 and 10-years gates and recovery of early predictive levels at 15 years. In contrast, the BTs showed a slight dip in their predictability between the 10 and 15-year periods for sexual

recidivism, but rather consistent increases in predictability over the 15-year follow-up for the two non-sexual crimes.

Comparisons between Child Molesters and Rapists

No differences emerged for the overall speed and frequency of sexual recidivism for the child molesters and rapists, with survival analysis estimated rates at 15 years being 24% and 20%, respectively for child molesters and rapists in the entire sample. In contrast, consistent with the hypothesis that rapists are criminologically more generalists than child molesters, rapists had significantly higher and faster rates of recidivism for both non-sexual crimes.

Differences in the temporal pattern of prediction emerged between the two groups. Rapists showed better initial predictability for sexual recidivism than child molesters, but this superiority dissipated over time. The cross-temporal analysis within each group showed predictability decreasing for the rapists between the 3rd and 10th year and increasing for the child molesters during this same period. Both remained consistent from the 10th to the 15th year. For most comparisons on non-sexual charges, rapists were better predicted across the follow-up than child molesters, with the cross-temporal predictions remaining constant for rapists and increasing for child molesters.

Comparisons of Different Crime Types

Although the results here appeared complex, the basic trend is that measures fashioned to predict sexual recidivism predict such better than they predict non-sexual crimes, and those created to predict criminality or psychopathy are slightly better at predicting subsequent non-sexual than sexual charges.

2. Meaningful Differences Among Modern Risk Assessment Instruments in their Level of Predictive Accuracy

The 3-year gate AUC coefficients for actuarials that were established empirically to predict sexual recidivism were approximately in the .670 to .700 range. They thus had relatively similar levels of predictive accuracy. These rates compare favorably with other follow-up studies (e.g., Craig et al., 2006; Harris et al., 2003; Langton, 2003). Across the extant studies no measure has arisen that has been consistently superior to others, with all measures varying substantially across studies. The SRA Need Assessment scale emerged as the highest predictor of sexual recidivism across the outcome periods, but it was significantly better than only the poorest measures.

Conceptual-mechanical (i.e., structured clinical guidelines) risk assessment instruments also had moderate though slightly lower levels of predictive accuracy. AUC coefficients were generally between .640 and .660 for the entire sample at the initial time gate assessed.

Detailed comparisons identified some actuarial subscales that fared more poorly than others in predicting sexual recidivism (e.g., the A-SOAP Dynamic scale and the RM2000 Violence Risk scale). Factor 1 of the PCL-R and the DSM-IV APD were also noteworthy as consistently poor predictors of sexual recidivism.

3. Can Cohesive, Meaningful Predictive Dimensions Be Identified in the Extant Risk Assessment Instruments?

The factors generated from the items of the extant actuarials were broadly consistent with those found by Barbaree et al. (in press). They indicate that five interpretable dimensions—Criminal Persistence, Sexual Persistence, Young & Single, Violent Sexual Assault, and Male Victim Choice--can be used to summarize a large number of static historical items and that the

resulting factor scores account for the predictive value of existing static risk assessment instruments. The first two factors consistently had the highest correlations with existing instruments, indicating the primary emphasis of the extant measures. The results of this factor analysis should be tested using confirmatory factor analysis in other samples, but they suggest that future prediction instruments can be explicitly constructed to measure the underlying dimensions. The results also open the possibility of integrating the results of existing instruments by identifying the underlying dimensions they assess and determining which operationalization of a factor yields the best reliability and validity.

4. Predictive Contribution of the SRA Need Framework

SRA Need Assessment yielded slightly higher AUCs than those of the best of the established risk assessment instruments. This is striking given the difficulty coders had in rating Need factors reliably from the MTC files. The SRA coding criteria require adjustments to improve reliability. Importantly, both static historical and Need variables made statistically significant contributions to prediction. An equally weighted combination of the two kinds of factors led to higher predictive accuracy, especially at the longer follow up (10 and 15 years).

5. Relation between Age and Sexual Recidivism

The present results suggest that age at discharge should not be introduced as a weighting factor for the actuarials. Strikingly, when age on index offense was controlled, increasing age on discharge was associated with increased rates of sexual recidivism., Those discharged after the age of 60, however, did have lower sexual recidivism rates, but the size of this sample was too small to make any firm recommendations. Nonetheless, the present results are consistent with adjusting expected recidivism rates down only for those discharged after the age of 60. The complexity of these results demand replication before too much weight is placed on them, but

they are certainly inconsistent with the hypothesis that merely holding someone in prison will enable them to age out of risk.

6. Determining Differential Predictors for Rapists and Child Molesters

Consistent with the finding that the survival curves for rapists and child molesters in predicting sexual recidivism did not differ, the addition to regression models of a dichotomous variable identifying the offender's status in these age-preference groups had no effect. In contrast, a strategy of focusing on subgroup specific measures that were either neglected or underutilized in extant actuarials and using these variables to predict subgroup-specific outcome yielded more promising results. Such measures were entered in discriminant function analyses (DFAs) to predict sexual recidivism at five time gates for rapists and child molesters separately. The earlier finding, when examining predictability over across numerous actuarial measures, that we might be able to predict more successfully the sexual recidivism of rapists in the short run and child molesters in the long run, was corroborated in these analyses. More importantly, five subtype-specific predictors for child molesters and two for rapists were identified. When these were entered into Cox Regressions to predict the hazard rate of sexual recidivism after the contribution of SRA Need Assessment had been removed, the five measures for child molesters accounted for significant additional predictive variance for the entire sample and the two measures for rapists accounted for significant additional predictive variance, but only for the BT sample, where they were more adequately measured. The measures included fixation on children, Paraphilias, male victim choice, Social Isolation, and Impulsivity for the child molesters, and pervasive anger and offense planning for the rapists.

The implications of these data for the current use of extant actuarials in the criminal justice system were discussed. Using the factor analytic results, the SRA Need results, and the

identification of differential age-preference subtype predictors, we are now working on the creation of a new actuarial.

Abstract

In a prior study using four outcome record sources we had followed 599 offenders who had been referred to the Massachusetts Treatment Center (MTC) for evaluation between 1959 and 1984. Of these 266 (Bridgewater Treatment [BT] sample) had been committed to MTC and subsequently released and 333 (Bridgewater Observation [BO] sample) had been determined not to be sexually dangerous and returned to finish their sentences. Of the 333 BO offenders 200 constituted a matched sample (on age at evaluation, marital status, and number of prior crimes) and 200 were randomly sampled from the entire sample BO population evaluated. The present study accessed the archival files for these offenders and coded them both on the modern actuarials that have been developed since 1998 and on a new experimental measure, the SRA Needs Assessment. Among the notable results were: (a) for the average predictability over all measures the BOs were better predicted than BTs, despite a significantly lower recidivism rate; (b) the cross-temporal pattern of prediction differed between rapists and child molesters, with the former being predicted better at shorter follow-up periods and the latter better at longer intervals; (c) all actuarials showed moderate reliability and predictive accuracy with few significant differences emerging in their direct comparisons with each other; (d) five factors accounted for all of the predictive variance in the extant actuarials; (e) although not significantly different from the best actuarials, nonetheless the SRA Needs Assessment consistently had the highest AUCs for the entire sample and for rapists and child molesters separately; (f) age was not found to constitute an important moderator for predicting outcome, and a complex relation among age at index offense, age at discharge, and outcome status emerged; and (g) promising additional subgroup specific predictors for child molesters and rapists were identified.

Evaluating and Improving Risk Assessment Schemes for Sexual Recidivism:
A Long-Term Follow-up of Convicted Sexual Offenders

INTRODUCTION

Need for Risk Assessment

Risk assessment serves multiple functions within the criminal justice system. Risk management options range from civil commitment within a secure setting at one extreme to multiple parole decisions at the other end. Whereas civil commitment incurs an annual cost well in excess of \$100,000 to restrict physically those with high potential for sexually coercive behavior, decisions about parole from prison involve substantially lower costs to achieve a similar goal. In both of these circumstances the accurate evaluation of the recidivism risk of convicted male adult sexual offenders would allow the concentration of risk management resources on those who pose the greatest risk. Within a prison setting, risk assessments may also inform decisions about the inmate's security level and about the amount and nature of the treatment services that are offered. Similarly, professional risk assessment may determine both the intensity of supervision and the nature of treatment services required of an offender under parole or probation supervision. Furthermore, the degree of involvement by police, mental health, and other community services in preventing future offending can be (and in the United Kingdom are) dependent on assessments of the degree of risk presented by the offender. Similarly, in some parts of the United States, community notification processes may depend on the level of assessed risk.

These various kinds of application of risk assessment involve different potential costs for all those involved. In some instances where risk is low the costs may be sufficiently low that simple administrative procedures may suffice. Even in such cases, however, more accurate risk

assessment can potentially save public money, lead to fewer sexual offenses, and reduce unnecessary hardships for individuals with a record of past offending. In other instances where the risk of recidivism is high, the potential costs are considerably higher for all parties. Because the offender, if not appropriately managed, may represent a real risk of serious and repeated offending, the public danger may be great. Here the decision to employ relatively drastic risk management procedures (for example, civil commitment, lifetime active GPS monitoring, extremely long prison sentences) involves very substantial commitments of public resources that require either increased taxation or the reduced funding of other programs. Moreover, they can impose considerable hardship and reduce the life-chances for individuals with a record of past offending. Thus, in such “high stakes” assessments the most accurate and justifiable risk assessments possible are required. Consequently, such “high stakes” risk assessments are often resolved in judicial or quasi-judicial forums.

History of Structured Risk Assessment

Until relatively recently professionals assessing the risk level of a convicted sexual offender typically employed what is now characterized as “unstructured clinical judgments.” At best assessors reviewed the offender’s file, interviewed him, perhaps spoke to others who knew him, and tried to form an integrated clinical picture of the offender. The final risk judgment was based on this global formulation and was guided neither by empirically informed theory or specific research findings.

An early review of assessments made in this way suggested that their predictive accuracy was only a little better than chance (Hanson & Bussière, 1998). Since that time considerable research effort has been devoted to developing more structured and accurate clinical procedures for assessing the recidivism risk of sexual offenders. The most successful approach has focused

on generating empirically derived, mechanical actuarial risk assessment instruments (Barbaree, Seto, Langton, & Peacock, 2001; Hanson & Morton-Bourgon, 2004; Langton, 2003).

Researchers typically developed these by selecting easily available, historical items that have been found to correlate with sexual recidivism in follow-up studies. The ability of various combinations of these items to predict sexual recidivism is first examined in a construction sample. On the basis of these preliminary analyses the final form of the scale is determined. The final version of the scale consists of rules both for scoring individual items and for combining item scores into an overall risk score. At this stage new samples are examined to assess the statistical properties of the scale. As cross-validation samples accumulate, more precise estimates of the recidivism rates associated with the instrument's various risk-bands emerge, and the scale's overall predictive accuracy is enhanced. Additionally, scale parameters may be found to covary in lawful ways, suggesting how the scale can be combined with other potential predictors. Finally, the applicability of the scale to increasingly diverse samples is examined.

As this body of research on a scale becomes more extensive clinicians are able to use the scale with increasing confidence, understanding both its strengths and limitations. Some empirically derived actuarial scales intended for use with sexual offenders have been repeatedly cross-validated. The most notable of these are the RRASOR (Hanson, 1997), the Static-99 (Hanson & Thornton, 2000), the Static-2002 (Hanson & Thornton, 2003), the SORAG (Quinsey, Harris, Rice, & Cormier, 1998), the MnSOST-R (Epperson, Kaul, Huot, Hesselton, Alexander, & Goldman, 1998), and the Risk Matrix 2000 (Thornton et al., 2003). Some of these instruments are used widely. Internationally, the Static-99 is probably the most extensively used instrument. It has been adopted for routine use by criminal justice agencies in places as diverse as Texas and Canada. The S scale from Risk Matrix 2000 has similarly been adopted as a standard risk

assessment instrument by criminal justice agencies in the United Kingdom. Meta-analyses (Hanson & Morton-Bourgon, 2004) have confirmed that empirically derived actuarial instruments typically achieve a real, but moderate level of predictive accuracy.

An alternative, but less researched, method of structuring clinical judgment has involved the creation of professional guidelines. These typically focus the assessor on factors deemed to be relevant for sexual recidivism on the basis of professional consensus, and they provide some guidance on how to score the items. The integration of the factors into an overall judgment may be left to the individual clinician's discretion. Such an approach allows clinicians to consider a wider range of factors than are included in the current generation of actuarial instruments. Additionally, it allows the clinician to take into account factors that may be relevant in the individual case, even if they are not generally relevant. In research the items of instruments of this kind are typically integrated according to some specified rule. In this case they become a mechanical assessment instrument. Their credibility still depends more on professional consensus about the items than on specific statistical studies. Importantly, however, when the items are combined mechanically we can accumulate empirical knowledge about the scale's properties in the same way that we gather data about an actuarial instrument.

Alternatively, researchers could study the predictive properties of actual clinicians integrating the items through their own idiosyncratic judgments. Unless a large and representative sample of clinicians were studied, however, the results of such an investigation would be hard to generalize. Moreover, given the track record of such clinically derived decisions (Grove, Zald, Lebow, Snitz, & Nelson, 2000), there is a high probability that clinicians' amalgamation of items would not be superior to a mechanical combination of the same items.

The most prominent example of this kind of instrument in sexual offender research is the SVR-20 (Boer, Hart, Kropp, & Webster, 1997). This instrument consists of 20 items, each specifying judgments about the degree of presence of a broad, clinically meaningful construct. This diverse group of constructs includes: features of the offender's history, sexual interests, personality, past offending, particular problems (with relationships, with substance abuse, with employment, with mental illness, etc.), and factors deemed relevant to his response to risk management strategies (e.g., attitudes about intervention, denial that he has a problem). Meta-analytic results (Hanson & Morton-Bourgon, 2004) suggest that a slightly lower, but broadly similar level of predictive accuracy is typically achieved with this method. The smaller number of studies involved means that predictive accuracy is known with less precision than is possible for the actuarial instruments. The SVR-20 itself has a relatively high average predictive accuracy, but with results varying widely between individual studies.

One of the criticisms of the existing actuarial instruments is that, although they allow us to discriminate offenders into broad groups that differ in their expected recidivism rates, they do not identify specific treatment targets. A different approach to assessment focuses on stable, but potentially changeable psychological characteristics that are empirically associated with and hypothetically predictive of repeated sexual offending. Because such constructs form natural targets for treatment, they have been referred to as criminogenic needs in the more general correctional literature. Instruments have been developed that combine static historical items with items indicating postulated criminogenic needs (e.g., Andrews & Bonta, 1995). Tools of this kind are sometimes called risk/need instruments. Criminogenic needs are also sometimes referred to as stable dynamic risk factors or as psychological risk factors.

Within the sexual offender field there are two notable, and rather similar, Risk/Needs instruments—STABLE and the Structured Risk Assessment (SRA) model. STABLE is a structured rating scale designed to be used with sexual offenders under supervision in the community. It is intended to be used with Static-99 so that the combination provides a risk/need assessment. STABLE was assembled on the basis of review of relevant research, theory, and consultation with other researchers. Thus, it falls in the professional guidelines tradition rather than in the empirically-derived actuarial tradition. Hanson (2006) presented preliminary results from a large scale trial of STABLE, but the study is still in progress and its results have not been published. Similarly, the SRA model (Thornton, 2002) proposes an integration of an established actuarial instrument (either Static-99 or Risk Matrix 2000) with a predefined group of needs. SRA includes a conceptual framework for identifying needs applicable to a specific offender, and a methodology for determining when additional constructs should be added to the needs framework. SRA leaves open the question of how the needs are best scored. Two operationalizations of it have, however, been widely used. Self-report questionnaires administered prior to treatment under non-adversarial conditions have been used in a number of studies (Thornton, 2002; Thornton & Beech, 2002; Beech, Fisher & Thornton, 2003; Craig, Thornton, Beech, & Browne, in press) and moderate levels of predictive accuracy have emerged, similar to those obtained using empirically derived actuarials. The second way of operationalizing SRA has been through structured clinical ratings. Her Majesty's Prison Service routinely uses an unpublished version of this system (SARN – Structured Assessment of Risk and Needs). Although SRA itself is not designed to be a mechanical instrument, particular operationalizations yield mechanical instruments. In general, SRA is better conceptualized as a

heuristic framework that can be used to guide the selection and organization of variables from any relevant data set.

There are a number of different ways to classify both the scales and the items they comprise. Hanson (2006) has recently proposed that structured instruments be divided into empirical-actuarial and conceptual-actuarial. Hanson used the word actuarial to refer to scales that have items with defined rules. Here we make a somewhat similar distinction between empirically derived actuarial scales and conceptual-mechanical instruments, with the latter referring to instruments composed of psychologically meaningful constructs selected on the basis of professional consensus, literature review, and theory. Additionally, the items used in scales are distinguished into static historical variables and needs (i.e., postulated stable dynamic risk factors). Whereas empirically-derived actuarial scales tend to comprise simple static historical variables (e.g., ever convicted of a sexual offense against a male victim; number of prior convictions for non-sexual offenses), conceptual-mechanical instruments may include many items targeting needs (e.g., sexual interest in children; impulsiveness) as well as other kinds of items, including psychologically meaningful, but fixed aspects of the offender's history (e.g., subject to serious child abuse, history of adolescent antisocial behavior).

These distinctions have some practical importance. Whereas needs directly identify treatment targets, the other kinds of items do not. In contrast, simple static historical items may be easier to determine objectively and cheaply (at least where good criminal records are available), so assessments based on them have both an economic advantage and may more easily be used in large-scale statistical research.

Issues Raised by Multiple Structured Risk Assessment Instruments

The development of multiple structured risk assessment instruments has raised the issue of what to do when their application yields divergent results. Individual evaluators adopt a range of strategies. Common ones include (a) adoption of a preferred instrument, (b) assigning the highest risk category obtained from a range of instruments, (c) assigning the lowest risk category obtained from a range of instruments, and (d) mapping instruments into some kind of dimensional framework.

The viability of the first strategy requires empirical evidence of the relative superiority of particular instruments for specific decisions. There have been very few studies that have compared the predictive accuracy of different instruments within the same sample. Harris et al. (2003) compared SORAG, RRASOR, and Static-99. They reported SORAG to be somewhat more predictive than the other instruments, though confidence intervals for the estimates of predictive accuracy overlapped. RRASOR was the least predictive. By contrast, Langton (2003, Table 2.8) found RRASOR to be the most predictive and SORAG to be the least predictive of the three implements, though again their confidence intervals overlapped.

Although Strategies b and c are commonly used, they are difficult to justify conceptually. Despite the inherent attractiveness of some kind of averaging procedure, Seto (2005) was unable to identify a scale integration strategy that improved on the predictive accuracy of the best individual scale. Of course, because the relative predictive potency of scales fluctuates across studies, this may be too high a standard to adopt for evaluating scale integration strategies. It is notable that averaging did as well or better than other algorithms. Unfortunately, there are too few studies that have compared multiple instruments in the same data set for this matter to be resolved empirically.

Strategy d depends on the development of a dimensional framework that subsumes the various assessment instruments. Roberts, Doren, and Thornton (2002) provided evidence that three dimensions contribute to actuarial risk assessment instruments: sexual deviance, antisociality, and youth. Barbaree, Langton, and Peacock (in press) factor analyzed the items from a more comprehensive set of instruments and identified six factors that they labeled Antisocial Behavior, Child Sexual Abuse, Persistence, Detached Predatory Behavior, Young and Single, and Male Victim(s). They reported that only three of the six factors (Child Sexual Abuse, Persistence, and Young and Single) contributed significantly to the prediction of sexual recidivism. If this finding were replicated, it would suggest that existing scales should be restructured to concentrate just on the dimensions with predictive potency. It is likely, however, that the optimum weighting of underlying dimensions will fluctuate somewhat among samples. Additionally, factor structures will need to be replicated before a determination of their average (cross-sample) predictive potency can be attempted.

Another approach to developing a theoretical framework focuses on integrating etiology and risk assessment (Beech & Ward, 2006). Because of the development of integrated theoretical models of sexual aggression with empirical support (Daversa & Knight, in press; Knight & Guay, 2006; Knight & Sims-Knight, 2003, 2004) we are at the point that a process-oriented approach to the assessment of stable dynamic factors can be undertaken and the assessment of risk can be woven into a life course perspective on the development and continuance of sexually aggressive behavior.

Improving Structured Risk Assessment

There are a number of possible routes to improving existing structured risk assessment instruments. If there were robust differences in predictive accuracy among instruments, we could

simply move towards focusing on the instruments with higher average predictive accuracy. Meta-analysis (Hanson & Morton-Bourgon, 2004) suggests, however, that the differences among extant instruments are small. Existing empirically derived actuarial instruments have broadly similar predictive accuracy, but each weights underlying dimensions differently (Roberts et al., 2002). A process-oriented strategy that identifies underlying core processes might be able to optimize the weighting of these core components. Such a strategy would have the advantage of conceptualizing risk from a theoretical perspective, rather than simply relying on the atheoretical aggregation of known predictors. This perspective could integrate well-established models from personality, psychopathology, criminology, and cognitive neuroscience to generate and test specific theory-driven hypotheses about core processes (i.e., dynamic traits) that lead to and maintain sexual aggression. Assessment of core traits and knowledge of their functioning within the nomological network of causal factors in sexual aggression would allow optimal weighting. Such a strategy could open up a range of more sophisticated measurements.

Barbaree, Blanchard, and Langton (2003) have argued that there is a significant effect of age at release on sexual recidivism rates that is not adequately accounted for in most actuarial assessment instruments. It is possible therefore that weighting age more heavily would allow improved predictive accuracy. It would be particularly important if the effect of age applied to aging in prison. If this were true the very long sentences now being applied to sexual offenders in the USA would produce a population that was much lower risk by the age of release.

Several studies have suggested that “need” as defined by the STABLE or the SRA contributes predictive weight that is independent of Static-99. This raises the possibility that need could be used to adjust the results of empirically derived actuarials and yield higher predictive accuracy.

Existing instruments are applied regardless of whether the sexual offender has offended against children or against adults. It is possible that the dimensions underlying existing empirically derived actuarial instruments differ in their predictive power depending on the offender's characteristic victim age preference. Predictions tailored on this basis might be more powerful. In their work developing the Static-2002, Hanson and Thornton (2003) presented some evidence against this, but the Static-2002 does not weight domains heavily that are important in child molestation, so this strategy is still viable.

In addition to considering subsamples of offenders and expanding the pool of predictors to include theoretically relevant variables not typically included in current actuarials, one can also apply more sophisticated scale construction techniques (Wright & Masters, 1982) and one can introduce non-linear analytic statistics to create improved algorithms for combining predictors (Bates & Watts, 1988; Breiman, Friedman, Olshen, & Stone, 1984; Huet et al., 1996; Lewis, 2000; Ripley, 1996; Seber & Wild, 1989), in an attempt to develop a more accurate actuarial instrument.

The Bridgewater Data Set

The Bridgewater data set provided a rare opportunity to investigate some of these issues. Follow-up data from multiple sources had been secured on a large sample (599 sexual offenders), who had between 2 and 28 years of potential on-street time post release. Thus, their follow-up time constituted the longest available of extant follow-up studies, and the sample size was reasonably large. Both offenders civilly committed as sexually dangerous and those evaluated, but not committed were among the sample, making it a diverse sample and allowing determination of differential prediction for offenders at varying risk. All offenders had had their archival records coded in an extensive dictionary, providing a wealth of additional assessment

variables and information. The files of those committed were quite extensive and detailed. The archival information of those evaluated but not committed varied, with comprehensiveness being inversely related to date of release. It was the task of the present project to code all of these offenders on a comprehensive set of the most popular actuarial instruments.

Issues Addressed in this Report

The present study used Bridgewater case files to code a wide range of modern risk assessment instruments. After coding was complete, the newly coded actuarials and additional scales created from the coding dictionary were used to predict recidivism data that had already been gathered. Several issues were addressed.

- (1) Overall, what level of reliability and predictive accuracy can be obtained, when modern risk assessment instruments are applied to a sample like the Massachusetts Treatment Center (MTC) offenders? The archival records were written prior to most of the research on sexual aggression by clinicians not privy to current scales and clinical knowledge. Can the actuarials be reliably coded on such records? The uniqueness and complexity of the MTC sample lends itself to multiple questions. First, how comparable are the predictive results for this somewhat different sample relative to prior studies. Second, because there are two distinct subsamples—those committed as sexually dangerous (the Bridgewater Treatment [BT] sample) and those judged not sexually dangerous and returned to complete their prison sentences (the Bridgewater Observation [BO] sample), we can examine whether the actuarials differ in their predictive potency for these two samples, which differ in their overall risk level. The sample can also be divided into rapists and child molesters and the ability of the actuarials to predict outcome for these subsamples can be examined. In addition to serious sexual crimes, we also coded both victim-

involved and victimless non-sexual crimes post release. We can examine the relative power of the actuarials in predicting these non-sexual offenses. Finally, the length of the outcome allows the examination of whether the actuarials have the same predictive potency for short- and longer-term outcomes.

- (2) Are there meaningful differences among modern risk assessment instruments in their level of predictive accuracy? Given the results obtained in previous research, it is likely that confidence intervals for the predictive accuracy of the different instruments will overlap. The size of the sample and the length of the follow-up period might, however, contribute to better differentiation among predictive algorithms and allow some differentiation of those measures that are differentially better or worse at predicting various kinds of criminal outcome. At the very least the present study will contribute important data to the meta-analyses of similar studies.
- (3) Can the static historical variables in modern risk assessment instruments be reconceptualize and refashioned into a small number of meaningful predictive dimensions? Several questions will be addressed here. Can meaningful dimensions be identified? Are they predictive? If combined do they yield prediction that is equal or superior to the scales from which the items come? Do the scales have predictive value when the static historical dimensions are controlled?
- (4) What level of predictive accuracy does a previously untested and newly developed measure, the SRA – Need Assessment attain? In addition, the study explored the extent to which the SRA-Need Assessment adds predictive value, when static historical variables are controlled, and determines the level of prediction obtained by combining these two kinds of variables.

- (5) Would increasing the weight given to age on discharge increase the predictive accuracy of existing assessment instruments? In addition to investigating the predictive value of age on release, the study also investigated the predictive value age at index offense, and the combination of these two variables. The present sample was unusual because the correlation between the two variables was low enough to make this possible.
- (6) Are the predictive measures significantly moderated by whether the offender was a child-molester or a rapist? To what extent can predictive accuracy be improved by tailoring assessment on the basis of the age of victims against whom the offender has previously offended?

Issues 2 to 6 speak to different ways in which risk assessments might be improved.

METHOD

Participants

The participants in this study were selected from two groups of sexually aggressive offenders evaluated at the Massachusetts Treatment Center for Sexually Dangerous Persons (MTC) in Bridgewater, Massachusetts between 1959 and 1984. MTC was established in 1959 under special legislation for the purpose of evaluating and treating individuals convicted of repetitive and/or aggressive sexual offenses. The legislation provided for a civil, day-to-life commitment by the court. Release was contingent on being found no longer "sexually dangerous."

Between 1959 and 1984 approximately 3600 state prison offenders were recommended for consideration as sexually dangerous to be evaluated at MTC. After a screening in which two psychiatrists visited each state prison, evaluated the prisoner, and wrote a recommendation, approximately 1450 of these offenders were transferred to MTC for a full evaluation. Of these

1450 offenders approximately 360 were committed for day to life (Bridgewater Treatment patients [BTs]) and 1090 were released back to their state prisons to finish their sentences (Bridgewater Observation patients [BOs]).

Between 1985 and 1987 we followed through multiple record sources all committed offenders ($n = 266$ of the 360 BTs) who had been released from MTC between 1959 and 1984) and two subsamples of 1090 BO offenders who had been evaluated at MTC between 1959 and 1984 and determined *not* to be sexually dangerous. The subsamples of BOs were selected by two different independently applied procedures--random and matched. The random sample was chosen by using a random number generator to select 200 offenders from the 1090 BOs. The matched sample was established by generating a subsample of 200 offenders from the 1090 who were closely matched on age at evaluation, marital status, and the number of prior crimes to the BT sample. Applying these two procedures independently, we selected 67 offenders who were in *both* the random and matched subsamples. Consequently, the *combined* BO sample had 333 offenders (67 in both random and matched subsamples and 133 offenders uniquely in each subsample).

By virtue of their commitment for “sexual dangerousness” the BT offenders should be at a higher risk for recidivism than the BOs, who were referred to MTC for extensive evaluation, but determined not to be “sexually dangerous,” and returned to prison to complete their sentences. Offenders in these BO groups should constitute a slightly higher risk group than generic sexual offenders. The reasons for referral to the initial group of 3600 offenders was, however, quite arbitrary, with particular districts of the state indiscriminately referring many offenders including a considerable number of offenders with simply “nuisance” or non-contact sexual offenses, other districts being more discriminating, and others referring no one. Therefore,

the BO group should be at lower risk than the BTs and more comparable to “contact” sexual offenders in general prison populations, but slightly more selective. The fact that approximately 40% of the combined BO sample had only one serious sexual charge and conviction attests to their lower risk.

The mean age of the entire sample at time of discharge either from MTC or from prison to the streets was 36.55 ($SD = 11.94$, range 17.44 – 78.99 years). They committed their first serious sexual offense at an average age of 25.04 ($SD = 8.77$, range 10 – 60 years). They committed their index offense, which brought them to MTC, at 30.95 ($SD = 10.74$, range 14.29 – 64.58 years). Figure 1 presents in ten-year intervals a histogram of the number of offenders who fell into each decade block for their age at the time of their first offense, and Figure 2 presents a similar histogram for their age at the time of the index sexual offense before their observation at MTC.

Table 1 presents some descriptive variables for the BTs and the *combined* BO group. As can be seen in the table, BOs were released at significant younger ages than BTs [$M = 35.5$ and 37.9 , respectively, $F(1, 560) = 5.26, p < .025$]. BOs also achieved significantly *lower* full scale IQs than the BTs [$M = 95.1$ and 99.6 , respectively, $F(1, 485) = 11.13, p < .001$]. Intelligence tests varied in the archival record from individually administered tests like the Wechsler Adult Intelligence Scale (the majority) or the Stanford-Binet Intelligence Scales to paper and pencil tests like the Shipley-Hartford. When multiple tests had been administered, we gave priority to the individually administered test. Prior to the index discharge BOs had significantly *more* nonsexual non-violent and sexual, non-contact offenses than BTs ($F(1, 485) = 3.98, p < .05$ and $F(1, 144) = 5.63, p < .025$), but significantly fewer serious sexual offenses ($F(1, 583) = 22.52, p < .001$). There was no difference in the percentage of Caucasian offenders in either group, $\chi^2 =$

0.32, *ns*. For 137 of the BOs the index sexual offense was the only serious sexual offense with which they had been charged in either adolescence or adulthood. For the 51 of the BTs the index serious sexual offense was the only sexual offense for which they had been charged.

For 504 of the offenders in the study there was sufficiently detailed information in their clinical files, that they could be categorized in either the Massachusetts Treatment Center Child Molester Typology; Version 3 (MTC:CM3; Knight, Carter, & Prentky, 1989) or Rapist Typology; Version 3 (MTC:R3; Knight & Prentky, 1990), two typological systems that require considerable detail about behavior, fantasies, and attitudes. Of the 504 offenders 248 met the criteria for classification as child molesters (all their victims were under the age of 16) and were classified in MTC:CM3, and 207 met the criteria for classification as rapists (all their victims were 16 or older) and were classified in MTC:R3. Of these 504 offenders 49 could not be classified in either typological system, because they had victims both under and over 16, and were therefore considered mixed offenders. A sexual offense was defined as any sexually motivated assault involving physical contact with the victim.

Table 2 presents some descriptive variables for these three age-preference groups. Rapists were discharged at a significantly younger age than the other two groups ($F(2, 468) = 19.05, p < .001$; Newman-Keuls $p < .05$). Rapists also had been convicted of fewer serious sexual crimes than either child molesters or mixed-age offenders ($F(2, 499) = 12.53, p < .001$; Newman-Keuls $p < .05$). Child molesters had a higher percentage of Caucasians than either rapists or mixed-age offenders ($\chi^2 = 10.74, p < .005$).

The final total follow-up sample that was analyzed comprised 566 offenders. Thirty-three offenders, both from the BT and BO samples, were dropped for a variety of reasons that included dying during the follow-up, discovery that after a BO was returned to prison, he was never

released until after the end of the follow-up period, and insufficient information in his archival file to rate the actuarials reliably (typically the earliest MTC archival files). In the analyses reported the n's change as a function of missing data.

Data Sources and Coding Procedures

Overview of Sources and Data Management

There were two sources of data for the proposed study. The first source was the offender's MTC clinical and criminal archival records. The second comprised four record sources that were accessed to obtain comprehensive follow-up data.

Two coding teams were created. Dr. Knight oversaw Team A, which comprised one data management coordinator, who also helped with some offender classification and DSM IV judgments, and one coder, who primarily did offender classifications. Team A was based primarily at Brandeis University. It had access to the complete case files and to recidivism data. This team reviewed each case file, edited it to remove any clues as to whether the offender had recidivated, assured its anonymity, and attached a case identity number. The case file was then scanned, converted into a PDF file, burned onto a CD, and mailed to Team B, which was based primarily at Sand Ridge Treatment Center. Dr. Thornton oversaw Team B, with Ms. Daniels helping to coordinate Team B coding. Upon receiving the cases, Dr. Thornton printed the file and assigned it to a coder (Ms. Daniels or one of three research assistants). The assigned coder scored a predefined list of items and entered the results into an Access database. Copies of the Access database were then periodically returned to both the Team A and Team B coordinators. Coders were instructed to return any case file that they believed gave them any clue as to whether the offender had recidivated. This applied in a small number of cases. These were then checked, the clue was removed, and the case assigned to a different coder.

In general Team A was responsible (a) purifying, redacting, and scanning detailed copies of offenders' files, (b) for classifying all BOs using both the MTC typologies and the DSM IV Conduct Disorder and Antisocial Personality Disorder categories, and (c) classifying a subset of BTs using the DSM IV Conduct Disorder and Antisocial Personality Disorder categories. Team B was responsible for coding all actuarials and the PCL-R on all offenders in the study, and for classifying all BTs using the DSM IV Conduct Disorder and Antisocial Personality Disorder categories. Coders were not privy to any follow-up data. Dr. Knight corrected all of the follow-up data, and wrote the program to concatenate the various outcome measures. When all of the predictive data were complete, he joined the follow-up data to the predictive scale file.

Predictive Variables

MTC Archival Files

For all 601 offenders these records included all information gathered during the 60-day evaluation at the MTC. This information came from a variety of sources, including school and employer reports, probation and parole records, psychiatric and medical reports, psychological testing, and clinical interviews coalesced at the time of their evaluation for commitment. For the committed sample these records additionally include all information subsequently added during the participant's commitment (quarterly treatment reports, ward behavior and work reports, and subsequent assessments for release consideration).

MTC Archival File Coding Procedures

During an earlier study the clinical records were copied, redacted of all identifying information for research purposes, and assigned a random research identification number. Three trained research assistants in Team B recoded these detailed redacted clinical and criminal files on the actuarial instruments briefly described below and summarized in Appendix 1. Some of the

BO files either did not contain sufficient information for coding or offenders were subsequently found either to have died or not to have been released to the streets after their MTC evaluation, so a total of 569 cases were coded on all actuarials. Varying numbers of offenders were dual coded on various instruments (see the n's in Table 3) to determine reliability. These dual coded files were randomly selected and assigned, and the coders were unaware of which files were being double coded. These assistants also rated all 569 offenders on the dynamic variables described in Appendix 2 and diagnosed the BT sample on the Conduct Disorder and Antisocial Personality Disorder criteria of DSM IV. Although for most of the offenders there was adequate information to make reasonable diagnostic judgments about these two DSM IV categories, there were typically not sufficient details to make independent diagnoses of the other personality disorders or of schizophrenia, which are required in the VRAG and SORAG. For these diagnostic judgments we accepted the diagnoses made by the original clinicians that were recorded in the files. Two additional assistants under the supervision of Dr. Knight categorized, where appropriate, all of the offenders in the BO sample into the MTC:R3 or MTC:CM3 types and diagnosed these offenders and 50 randomly selected offenders from the BT sample on the Conduct Disorder and Antisocial Personality Disorder criteria of DSM IV. These latter diagnoses served to assess reliability. None of the research assistants had access to or were aware of the follow-up data. To make this possible the clinical files made available to coders were edited to ensure that only they included information about events that occurred prior to the start of the follow-up period. Thus they had access to the offenders' prior criminal history, but not to recidivism information. Not even the PI linked the data sources until all the files were coded and the predictive data had been entered into the computer database. Using Access we created computerized rating forms, so that all item and scale judgments and all notes taken on each

offender *were* directly typed into a desktop or laptop computer and automatically added to the database.

Actuarial Coding

Seven mechanical actuarials were rated from these files. One structured clinical instrument (SVR-20) was adapted for mechanical application. Finally, the PCL-R was rated for all offenders. These are described in the following sections and their items are presented in Appendix 1. The authors of the MnSOST-R provided training. In addition to instruction, the coders scored sample cases and received feedback on scoring errors. Coders emailed in scoring queries to the Team B coordinator and these were summarized and relayed to the authors of the various instruments. The latter invariably replied helpfully and their guidance was passed on to the coders.

Violence Risk Appraisal Guide (VRAG) and Sex Offender Risk Appraisal Guide (SORAG). The VRAG (Harris, Rice, & Quinsey, 1993), which was developed to assess risk for violent recidivism (including sexual offenses involving physical contact with the victim), contains 12 items (see Appendix 1). Item weights were calculated using the empirical relation between the predictor and violent recidivism in the development sample. The total VRAG scores, which can range from -26 to +38, allow assignment of an offender to one of nine risk categories, ranging from 1 (lowest risk) to 9 (highest risk). Despite the absence in the MTC records of the requisite phallometric assessment to diagnose sexual deviance on the SORAG (Quinsey et al., 1998), nonetheless this adaptation of the VRAG was also rated (see Appendix 1).

Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR). Using the results of a meta-analysis of 23,972 offenders (Hanson & Bussière, 1998), Hanson (1997) selected variables with a minimum correlation of .10 with sexual recidivism and developed a brief actuarial scale

comprising the four items that yielded the best independent predictors of sexual recidivism (see Appendix 1 for the items). An offender's total scores can range from 0 to 6. Despite its brevity the scale has yielded ROC curves comparable to more extensive scales (Barbaree et al., 2001; Langdon, 2003).

Static-99 and Static-2002. The ten items in the Static-99 (Hanson & Thornton, 1999, 2000; see Appendix 1) were selected from the non-redundant items in two pre-existing risk scales: the RRASOR and Thornton's Structured Anchored Clinical Judgment scale (SAC-J; Grubin, 1998). The Static-2002 is a recent revision of the Static-99, which is intended to improve the definition and rating ease of the Static-99 and to broaden the scale with additional non-redundant items (Hanson & Thornton, 2003). It contains the 17 items listed in Appendix 1.

Minnesota Sex Offender Screening Tool – Revised (MnSOST-R). The MnSOST-R (Epperson et al., 1998) was developed to be a screening tool that relied on information present in typical clinical files. The scale comprises 16 items, 12 coded using historical information and four coded using information about the offender's index offense (see Appendix 1). Item weights for the MnSOST-R were derived from multiple regression analyses (Nuffield, 1982) using individual items to predict sexual re-offense outcome in the development sample (see Epperson et al., 1998). Total scores on the MnSOST-R can range from -14 to +30. On the basis of these scores offenders can be assigned to either three or six risk levels (Epperson, 2000).

Adult Sex Offender Assessment Protocol (A-SOAP). Prentky and Righthand (2003) developed the A-SOAP as a parallel adult actuarial instrument for their Juvenile Sex Offender Assessment Protocol (J-SOAP, Prentky, Harris, Frizzel, & Righthand, 2000). The A-SOAP comprises 24 items (see Appendix 1) that have demonstrated an association with sexual recidivism in four domains—Sexual Drive/Preoccupation (8 items), Antisocial Behavior (7

items), Intervention (7 items), and Community Stability (4 items). It was possible for us to rate the third domain for many of the BO cases as well as for the BT cases, but the fourth domain could only be examined on the small portion of the sample released through the MTC Authorized Absence Program. All items were rated as either ‘not present’ (0), ‘possibly or partially present’ (1), or ‘present’ (2). One can analyze the total score (ranging from 0-48) and scores on each domain scale.

Risk Matrix 2000 (RM2000). The RM2000 (Thornton, 2002) is a statistically derived risk classification process intended for males ages 18 and older, who have been convicted of a sex offense. It uses simple factual information about offenders’ past history to divide them into categories that differ substantially in their rates of reconviction for sexual or other violent offenses. RM2000 consists of 3 scales: (a) RM2000/S is a prediction scale for sexual offending; (b) RM2000/V is a prediction scale for non-sexual violence engaged in by sex offenders; (c) RM2000/C is a combination of the first two scales and predicts sexual or other violence. The items for each scale are presented in Appendix 1.

Sexual Violence Risk–20 (SVR-20). Although the SVR-20 (Boer, Hart, Kropp, & Webster, 1997; Douglas, Ogloff, Nicholls, & Grant, 1999; Douglas & Webster, 1999) was developed as a structured clinical risk assessment scheme meant to guide clinical decision rather than provide an actuarial score, we used it the way that Langton (2003) did, assigning unit weights to its items and calculating a summative score. The SVR-20 comprises 20 items (see Appendix 1) that have demonstrated an association with sexual recidivism and/or that appear to be clinically relevant to sexual recidivism in three domains—psychological adjustment (11 items), sexual offenses (7 items), and future plans (2). It was possible for us to rate the last domain on the entire sample. All

items are rated as either ‘not present’ (0), ‘possibly or partially present’ (1), or ‘present’ (2). One can analyze the total score (ranging from 0-40) and scores on each domain scale.

SRA Need Variables. An outline version of the current SRA Need framework is presented in Table 3, and more details are provided in Appendix 1. A preliminary review of a small sample of case files was undertaken to determine which of the Need factors could be scored with reasonable validity from the available information in the MTC files. Factors selected for coding in the current data set are marked with a star. It should be noted that the primary exclusion is the Distorted Attitudes domain. This decision reflects the remarkable naivety about distorted attitudes displayed in the clinical files, which were written before the importance of cognitive distortions had been established.

Rating scales applicable to the current data were written for these factors. Each factor was scored 0 (not applicable), 1 (partially applicable), or 2 (generally applicable). Because PCL-R factor two is a prototypical marker for the Lifestyle Impulsiveness factor, an additional rating scale was not written for this factor, but scores were recoded into the same three values using cutting lines that coded it as 0 for raw scores below the sample mean, coded it as 2 for raw scores more than one standard deviation above the sample mean, and coded “1” if raw scores were between these two levels.

Once the data had been collected, items from the other scales that were conceptually related to the SRA Need factors were identified. Correlations between the items written as SRA factor markers and these conceptually related items were calculated. Where an appropriate pattern of correlations (high between conceptually related factors; lower with other factors) was obtained, the conceptually related item was included in the total SRA Need score. All these additional variables were already coded 0, 1, or 2 so their coding was left unchanged.

Because SRA assumes that the different Need domains should be about equally weighted in computing an overall Need Index, a check was run to see that domain scores correlated about equally with the SRA Need score. These correlations were all around .70 indicating that this condition was met. The correlation between Need scores from independent raters was .66 ($n = 89$; $p < .001$).

Psychopathy Checklist-Revised (PCL-R). The Psychopathy Checklist – Revised (PCL-R, Hare, 1991) is a structured clinical assessment instrument developed to assess psychopathic personality traits. It comprises 20 items that are scored on the basis of a file review (see Appendix 2). All items are rated as either ‘not present’ (0), ‘maybe or in some respects’ (1), or ‘present’ (2). One can analyze the total score (ranging from 0-40), or one can analyze scores on each of two hierarchical factors (Arrogant and Deceitful Personality/Emotional Detachment and Impulsivity/Antisocial Behavior [Cooke & Michie, 2001; Hare & Neumann, 2006]). Although the recommended total score cut-off for classifying an offender as psychopathic is 30, 25 has been found to be effective for predicting future sexual offending (Quinsey et al., 1998).

MTC Coding Dictionary. Also available to us were all of the data from the MTC Coding Dictionary (see Appendix 3 for summary of the scales currently examined) that have previously been coded by two raters and entered into the computer. Although there is some overlap between the items in the dictionary and some of the items on the actuarials, they are sufficiently different and there would be enough missing items to warrant recoding. Moreover, Barbaree et al. (2001) found that when they attempted to rate actuarials by adapting existing computer codings the actuarial scales performed less well than when the items were directly coded from the clinical records. Using factor analytic strategies and rational scale construction, Dr. Knight created a

number of scales from the MTC Coding Dictionary that were theoretically related to recidivism. Some of these will be used in the analyses reported below.

Follow-up Data Sources and Procedures

Follow-up Sources

The second data source is the criminal outcome records of the offenders. Recognizing the problematic nature of the collection of information on the post-release criminal behavior of ex-offenders (Jenkins, Barton, de Valera, DeVine, Witherspoon, & Muller, 1972; Jenkins, DeVine, de Valera, Muller, Nichols, & Ray, 1975), we considered it essential for data enrichment and cross-validation purposes to include as many official data sources as were available. Thus, we initially selected five criminal record sources: the Massachusetts Board of Probation records, the Massachusetts Parole Board records, the Massachusetts Department of Public Safety records, the Massachusetts Treatment Center Authorized Absence Program records, and the Federal Bureau of Investigation (FBI) records. The Department of Public Safety records was subsequently dropped as a data collection source after it became apparent that the available information was identical to that found in both the FBI records and the Department of Correction records. The four remaining sources (Probation, Parole, MTC files, and FBI) were highly reliable (i.e., each of these four sources kept records on all of the offenders in our sample). Although they provided somewhat redundant information, the multiple sources allowed for crosschecking of information. Moreover, the FBI records allowed us to identify out-of-state crimes that might be missing in the state records. The major limitation of these data was that they only identified offenders who came into contact with the law. Thus, if an offender did not commit any new offenses or was not apprehended for ones he did commit, no information about him appeared in the records.

Massachusetts Board of Probation. The central office of the Massachusetts Board of Probation (Office of Commissioner of Probation) keeps a continuing record of every individual processed through the judicial system. The information available consisted of all charges and the consequent dispositions. This source was reliable for both misdemeanour and felony charges for offenses committed within the State of Massachusetts.

Massachusetts Parole Board. The Massachusetts Parole Board keeps records on all individuals incarcerated in correctional facilities and released on parole. The parole board provided information on most, but not all of the offenders in this sample. The board does not keep records on (a) inmates who have served their total sentences, (b) inmates at the Treatment Center who have been found not to be sexually dangerous and released from their indeterminate commitments through court order, or (c) inmates who have been placed on court-ordered release programs, regardless of any sentenced time that had not been served.

Massachusetts Treatment Center Authorized Absence Program. Release from the Treatment Center may be outright, under Section 9 of Chapter 123A, if an individual is declared no longer sexually dangerous and has resided at the Treatment Center for a period of time equal to or exceeding his prison sentence. An individual may also be "gradually released" from the Treatment Center through the Authorized Absence Program. If an individual is placed on a program, the Treatment Center keeps the individual's file open. There are monthly status reports reviewing the inmate's program, and case managers submit notes on his activities and community adjustment.

Federal Bureau of Investigation. The FBI keeps an ongoing record of all individuals convicted of felonies. Their records contain charges and dispositions for offenses committed throughout the nation. Felony charges are reported routinely, but misdemeanor charges are

reported less consistently. Although the FBI records were less reliable than those from the Board of Probation, they provided a crosscheck of within-state information and included information on offenses committed outside of Massachusetts, which were not available in the Massachusetts records. Consequently, the outcome dataset includes both in-state and cross- or out-of-state offenses.

Follow-up Procedures

In 1987 and 1988 two trained research assistants had accessed all of the above outcome sources and recorded for each offender all subsequent information that was available on him between release from MTC and March 1, 1986. The sources were integrated into a timeline that included every recorded charge arranged chronologically within the follow-up period. For each charge that was coded, the date, description of the charge, the disposition, and the state in which it occurred were specified. During the follow-up period, we classified criminal offenses by generating a list of all possible criminal charges using a Commission of Probation handbook and an FBI handbook. In addition, coders added to the original list when they encountered charges that had not been initially included. The final list included a total of 172 criminal charges, of which a total of 78 different charges were coded (67 different charges for the rapists and 57 different charges for the child molesters). These offenses were divided into four global categories:

1. Serious sexual offenses, which consisted of 15 sexual charges that involved physical contact with a victim (e.g., carnal abuse, accosting, unnatural acts, indecent assault, assault with intent to rape, rape, sodomy, statutory rape, incest);

2. Nonsexual victim-involved offenses, which included 13 nonsexual offenses that required a direct encounter with a victim (e.g., assault, assault and battery, assault with intent to kill, assault with intent to rob, manslaughter, murder, robbery, armed robbery);
3. Victimless offenses, which consisted of 22 nonsexual charges that did not involve direct contact with a victim (e.g., larceny, trespassing, disorderly conduct, forgery, drunkenness, breaking and entering, malicious destruction of property, conspiracy, and possession of a dangerous weapon); and
4. Nuisance sexual crimes, which included all sexual crimes that did not involve contact with the victim (e.g., exhibitionism, gross and lewd behavior).

The data were coded so that we could examine outcome by charge, conviction, and reincarceration rates. It was our intention to define re-offense as broadly as possible. Of these three ways of measuring re-offense, however, we have found charges to be the most reliable index of the nature of re-offense. Within the Commonwealth of Massachusetts, charge is a more inclusive category than arrest (i.e., you may be charged without being arrested). The original charge is, moreover, substantially less likely to have been reduced to a lesser, nonsexual offense as a result of plea-bargaining. Only the analyses of criminal charges during the outcome period will be discussed in this report because the conviction and charge analyses were very similar. In general charges were slightly better predicted than convictions, but there were no differences between charges and convictions in the overall pattern or the contribution of predictors.

In this report we will focus mostly on the prediction of serious sexual offenses, but we will also provide some summative comparisons with nonsexual victim-involved and victimless offenses. We chose to concentrate on recidivism for serious sexual offenses and to exclude for the sexual recidivism analyses nuisance sexual offenses because: (a) in most important

applications it is serious sexual recidivism that is of public concern, and (b) including non-contact recidivism events in the sexual recidivism rate reduces the applicability of the results and leads to concern that scales may only be predicting less severe forms of sexual offending. It is true that some non-contact charges may reflect precursors to more serious sexual offenses, and there may be instances where plea-bargaining resulted in a contact offense being charged as a non-contact offense. Nevertheless, this is only likely to be true for a minority of non-contact offenses, so we did not feel that treating non-contact charges as if they typically represent contact offenses was justified.

Data Analyses

Several analytic techniques were applied to address the issues outlined at the end of the Introduction. These included Receiver Operating Characteristic (ROC) curve analyses, survival analyses, Cox Regression analyses, and discriminant function analyses (DFAs). We briefly describe each analysis type in turn and the computations that created the dependent, outcome measures used for each.

ROC Analyses

In ROC analyses the true-positive probability (sensitivity) of a prediction is plotted against the false-positive probability (1 minus the specificity) (Swets, Dawes, & Monahan, 2000). The area under the ROC curve (AUC) measures the predictive accuracy of the particular actuarial instrument. Because AUC values, unlike other indices, have the advantage of being relatively immune to selection ratios or base rates in the sample (Swets, 1986), they represent the best index of accuracy for relatively low base rate phenomena like sexual reoffending (Mossman, 1994; Rice & Harris, 1995). AUC values range from 0 (worse than chance prediction) to 1 (perfect prediction), with .5 representing chance level prediction. The AUC

value represents the probability that a randomly selected individual in the sample who re-offends will have a higher score on a given risk assessment instrument than a randomly selected individual who does not re-offend. Direct comparisons between AUC values for various instruments will be assessed using ROCKIT Version 0.9.1b (Metz, 1998), and SPSS, Version 13. The predictive efficacy of the various risk instruments will be compared at different outcome intervals, for predicting different crime types, and for predicting for different sex offender groups.

The dependent measures in these analyses were dichotomous items indicating non-recidivate (0) or recidivate (1) for a particular crime type within a particular time gate (3, 5, 7, 10, or 15 years). Only offenders who would have had the potential to be on the streets for the allotted time (e.g., for 3-year gate, they had to be released on or before February 28, 1983), and whose on-street time was not foreshortened by re-incarceration for another crime type (e.g., when considering serious sexual crimes for the 3-year gate, the offender was not re-incarcerated for another type of crime within the three year block) were considered in each time gate measure. This method of defining recidivism more nearly approximates recidivism rates for the defined time periods. It does, however, introduce a subtle restriction in the sample. For example, when calculating rates of serious sexual recidivism at various time gates, it under-represents offenders who are at high risk for other kinds of offending that lead to their imprisonment. In particular, offenders who recidivated more rapidly on non-sexual crimes may be under-represented in the analyses. If, however, they remained in the sample, they would have been counted as non-recidivists for sexual crimes for longer periods than they were actually on the street.

Survival Analysis Techniques

Survival analyses are a collection of methods for analyzing categorical outcomes that take into account not only whether members of each group commit subsequent crimes, but also the length of time between release and criminal activity, or the speed of reoffending (Harris, Kaylan, & Maltz, 1981; Schmidt & Witte, 1988). They conceptualize recidivism as a failure rate (FR) that takes into account the amount of time each offender has been on the street and able to re-offend. These methods have the advantage of allowing one to include in a single analysis all participants, both those who were followed for the entire follow-up period, and those that were followed for shorter durations. That is, they are flexible enough to consider incomplete observations, known as "censored" observations. Thus, they provide a statistical summary of all cases regardless of the length of time each was followed. They make the critical assumption that when one includes offenders who have not been followed for the maximum duration of the study, the shorter duration of their follow-up is not related either to something about their group assignment or to the outcome. The "censoring" of the offenders was a function of their release date and either the endpoint of our follow-up or their premature removal for another crime type (i.e., when considering serious sexual recidivism, reincarceration for another type of crime censored that offender at the time of reincarceration).

The survival curve represents the cumulative proportion of offenders who have not been charged with a crime by a particular point in time after release to the streets. The estimate of the curve is known as the Kaplan-Meier product limit or the nonparametric estimate of the survival curve. In the simplest analyses we will use the Mantel-Cox test or the logrank test, which yields a one-degree chi-square, to test the equality of survival curves for various risk groups identified by actuarials and for various subtypes of offenders (e.g., rapists versus child molesters).

Cox Regression Analyses

The Cox regression survival model applies regression methodology to survival data, allowing the testing of the relation of multiple variables and their interactions to risk of recidivism at a particular time (Hair & Black, 2002; Wright, 2000). The application provided an analytic framework to examine the relation between both continuous and categorical variables and survival and to address several important questions, including whether theoretically driven predictor models added explanatory variance to the best of the current actuarials, and whether different sets of variables enhanced prediction for rapists and child molesters. The dependent measures for the Cox Regression were the same as those used in the survival analyses.

Discriminant Function Analyses

The DFAs provided an alternative strategy to examining similar issues that were addressed by the Cox Regression analyses. Here the logistic group outcome of the time gates constituted the dependent variable rather than the overall hazard function.

RESULTS

The presentation of the results is organized using the six issues that were described in the Introduction.

1. Reliability and Predictive Potency of Actuarials Applied to the MTC Sample

Reliability of the Actuarials

Table 4 presents the interrater reliabilities for the total scores of the actuarials that were coded in the study. As we indicated in the Method section, when possible cases were deliberately assigned to more than one rater to allow the assessment of interrater reliabilities. Assignment of second codings was random, and raters never knew which cases had been or would be coded by

a second rater. When two coders scored cases, their resulting scores on the instruments were averaged, and the average score was used in subsequent analyses.

Table 4 shows the correlations between independent raters' scores on the various instruments together with the means and standard deviations for the various scales, the Ns rated by one or both raters. All the inter-rater correlations were significant at or beyond $p < .001$. Although some of the reliabilities were slightly lower than those reported in comparable studies (e.g., Langton, 2003), with the exception of the SVR-20, the differences are within a couple of hundreds. Moreover, only the SVR-20 had an interrater reliability below .70. The rest were acceptably high.

It is notable that the mean Static-99 score falls in the moderate-high risk band, reflecting the relative infrequency of low risk offenders in the sample. The mean PCL-R score is surprisingly low. This was due in part to the absence in many case files of the kind of information that would allow a confident determination of whether scores of 1 vs. 2 were applicable. Coders were advised to score 2 only when they could find positive evidence to justify such a score. The slightly lower interrater reliabilities than are typically reported in similar archival studies may to some degree reflect the quality of certain information in the files. As we indicated earlier, the records were written at a time that preceded not only the modern actuarials, but also the modern conceptualization of psychopathy and even the initial version of the PCL, so evaluators, who wrote the original reports, were not sensitized to the criteria that make up these scales. [Consequently, the reduction in clear file information on some complex psychological constructs may have lowered the reliability on such constructs, when judgments on them were required in various actuarials.](#)

Overall Predictive Accuracy in the MTC Sample

Table 5 shows ROC AUCs for the actuarial risk assessment instruments included in the study. Both AUCs and their confidence intervals (CI) are presented for the 3, 10, and 15-year fixed follow-up periods. The number of asterisks indicates the level of significance of the difference of each AUC from chance.

In general, the AUC levels for the commonly researched actuarials are comparable to (e.g., Langton, 2003) or better than (e.g., Craig, Beech, & Browne, 2006) other *independent* assessments at similar follow-up periods. As has been previously noted (e.g., Barbaree, 2003), non-independent evaluations by authors of instruments often exceed the values reported by independent researchers. The follow-up AUCs presented in Table 5 exemplify this comparability with previous research. The values presented for the Static-99 are .713 and .684, for 3 and 10-year follow-up periods, respectively. Langton reported an AUC of .63 for sexual crimes at 5.6 years follow-up for the Static-99. The 5-year AUC in the present study was .698 (CI = .639-.757). In a subsequent analysis of the 3-year follow-up, using only files on which all the actuarials he coded had complete data (only the most complete files were used), Langton reported an AUC of .74, clearly within the CI of our comparable follow-up time (.650-.777). Craig et al. (2006) had somewhat less success with the Static-99, reporting 5 and 10-year AUCs for serious sexual offenses of .59 and .52, respectively. Both of these were below the lower confidence intervals for the appropriate comparison time in the present study, but the 5-year follow-up AUC was within the lower end of Langton's CI range (.56-.71) for a similar period.

*Analysis Strategy for Comparing Predictive Potency
Across Subsamples and Temporal Gates*

Tables similar to Table 5 were calculated separately for BOs and BTs and for rapists and child molesters. For the entire sample and for each subsample, separate tables were generated for serious sexual, non-sexual victim-involved, and non-sexual victimless crimes. All of these tables are available from the authors upon request. To summarize the mass of data around the issues of the differential predictive potency of all measures across samples, crime types, and follow-up time gates, we turned the obtained AUCs into independent measures and calculated paired *t-tests* for all group and temporal comparisons of interest to assess predictive trends over the domains of interest. Analogous to the increased reliability that is attained when items are added to a scale (Nunnally, 1978), using multiple predictive measures together reduces error variance and increases the sensitivity for detecting trends in prediction over time and samples. Because all actuarials are highly correlated (Langton, 2003) and often cover the same predictive domains from slightly different perspectives, these analyses violated the *t-test* assumption of the independence of observations. To minimize this problem the analyses were repeated twice, once using only subscales less likely to overlap and a second time using only total scores. The results were comparable for all analyses, so we will present the analyses calculated on the 27 measures reported in Table 5. The method provides an exploratory assessment and summative description of trends within these data. Because the increased sensitivity in this method of analysis identifies significant differences that have small effect sizes, the results should be interpreted simply as descriptions of interesting trends in the predictive data.

*AUCs within the Entire Sample**Comparative Prediction at Various Time Gates*

The means presented at the bottom of Table 5 provide an example of the analytic strategy described above. As can be seen in the bottom rows of the table, the overall mean AUC across measures for the entire sample for each time gate were .640, .643, and .641, respectively for the 3, 10, and 15-year periods. The paired *t-test* analyses of time gates revealed that for sexual recidivism there were no overall differences in predictive potency across the three time periods, $t(26) = .696, .107, \text{ and } -.404$ for the 3 to 10-year, 3 to 15-year, and 10 to 15-year comparisons, respectively.

Table 6 presents the mean AUC values and standard deviations of the measures for the entire MTC sample and for four subsamples for different crime types for three follow-up gates—3, 10, and 15-years. The top data line for the serious sexual charges for the entire sample repeats the last two lines of Table 5, described in the prior paragraph. As can be seen in the table, in contrast with sexual recidivism, the mean AUCs for non-sexual violent recidivism for the entire sample consistently improved over time. At every time block the entire sample achieved higher average AUCs than the prior time block, and all three comparisons reached significance, $t(26) = 3.16, p < .01$ for the 3 and 10-year comparison, $t(26) = 6.76, p < .001$ for the 3 and 15-year comparison, and $t(26) = 4.93, p < .001$ for the 10 and 15-year comparison (see Table 6).

For non-sexual victimless charges no change in predictability occurred between 3 and 10 years, $t(26) = -.720$, but the 15-year AUCs were significantly higher than both the 3-year, $t(26) = 5.79, p < .001$, and the 10-year mean AUCs, $t(26) = 11.97, p < .001$.

Comparisons Between the BOs and BTs in Recidivism and Predictability

BO and BT Comparisons in Recidivism

Comparative Kaplan-Meier survival analyses between the BTs and the entire BO sample revealed that the former group recidivated at a faster and higher rate than the latter group only for serious sexual crimes, $LR(1) = 26.62, p < .001$ (see Figure 3). For non-sexual victim-involved and victimless crimes, there were no differences in survival curves $LR(1) = 1.25$ and 1.63 , respectively (see Figures 4 and 5). If one considered only the randomly selected BO subsample (the better estimate of all BOs evaluated at MTC than total BO sample in the present study, which also included a matched subsample), the only change that emerged was for the non-sexual victim-involved charges. For these charges the BTs also recidivated at a faster and higher rate than the randomly selected BOs, $LR(1) = 4.97, p < .05$ (see Figure 6). Whereas the recidivism rates for serious sexual charges for the BTs using the 3, 10, and 15-year gating criteria were 22% (45 of 206), 35% (53 of 151), and 35% (22 of 63), respectively, the *projected* estimates from the survival curves for the same time periods were 21% (52 of 246), 30% (74 of 246), and 35% (86 of 246). In contrast, the recidivism rates for serious sexual charges for the randomly selected BOs using the 3, 10, and 15-year gating criteria were 3.2% (5 of 154), 7.4% (9 of 122), and 10.9% (10 of 92), with the projected survival analyses estimates for the same time gates being, 2% (3 of 171), 8% (14 of 171), and 12.5% (21 of 171).

BO and BT Comparisons in Predictability

The comparisons on the overall prediction of serious sexual charges for the BOs and BTs were clear, consistent, and informative. Using the *t-test* strategy described above, we compared the mean sexual charge AUCs of the BTs and BOs across the 27 predictors for the 3, 10, and 15-year time gates. At every time block the BOs achieved higher average AUCs than the BTs, and

two of the three comparisons reached significance, $t(26) = 2.97, p < .01$ for the 3-year comparison, $t(26) = 2.00, p = .056$ for the 10-year comparison, and $t(26) = 5.75, p < .001$ for the 15-year comparison (see Table 6). The average AUCs for the entire sample were also significantly higher than those for the BTs at every time gate tested, $t(26) = 10.65, p < .001$ for the 3-year comparison, $t(26) = 7.84, p < .001$ for the 10-year comparison, and $t(26) = 9.24, p < .001$ for the 15-year comparison. The average AUCs for the entire sample were significantly higher than the AUCs for the BOs *only* at the 10-year period, $t(26) = 3.32, p < .01$. It has been argued that the predictability of sexual recidivism should increase as the predictive target increases (Barbaree, 1997). That is, as the percentage of recidivating offenders increases, moving from a low base rate frequency toward 50%, the ability of an actuarial to improve accuracy over simply using the base rates increases (Meehl & Rosen, 1954). The present data suggest that for determining predictive accuracy perhaps the increase in the range of risk that was present in the BOs relative to the BTs may have been more important than simply the size of the predictive target, which was greater in the BTs than the BOs..

Relative prediction of non-sexual crimes for BTs and BOs was not as consistent. For victim-involved offenses, the BOs had higher mean AUCs than BTs at the 3-year gate, $t(26) = 5.21, p < .001$, but no differences emerged at the 10 and 15-year follow-ups. AUCs for the entire sample were higher than those for the BTs at the 3-year, $t(26) = 8.97, p < .001$, and 10-year gates, $t(26) = 3.00, p < .01$, but not at the 15-year gate, $t(26) = .895, ns$. For the BOs victim-involved charges were predicted less well than for the entire sample only at the 15-year follow-up, $t(26) = 3.47, p < .01$. For victimless crimes BOs showed better predictability than the BTs at the 3, $t(26) = 7.32, p < .01$ and 15-year follow-up blocks, $t(26) = 2.30, p < .05$, but not at the 10-year gate, $t(26) = .693, ns$. BTs had significantly lower AUCs than the full sample at the 3, $t(26)$

= 8.97, $p < .001$, and 15 year gates, $t(26) = 3.18$, $p < .01$, but not the 10-year follow-up, $t(26) = .80$, *ns*. BOs differed from the entire sample only at the 3-year block, where they surprisingly yielded *superior* AUCs, $t(26) = 9.37$, $p < .001$.

The general group comparative trend across all crime types was that the entire sample yielded the best AUCs, followed closely by the BOs. The BTs, who showed the largest predictive target only on the sexual charges, were nonetheless comparatively more difficult to predict across crime types than the BOs and the entire sample.

BO and BT Comparisons of Predictability across Outcome Time

Although it might simply be a function of the large number of comparisons that we calculated, an interesting differential temporal pattern emerged for the BOs and BTs in the accuracy of the actuarials for sexual recidivism. As in the overall sample, the accuracy of the actuarials for predicting sexual recidivism for the BOs appeared to remain relatively constant over the follow-up period. There was a non-significant dip in predictability for the BOs from the 3 ($M = .624$, $SD = .069$) to the 10-year period ($M = .611$, $SD = .057$), but the AUCs recovered at 15-years ($M = .630$, $SD = .050$), with the 15-year gate AUCs posting a slight superiority to the 10-year AUCs, $t(26) = -2.56$, $p < .05$. In contrast, the predictability of the BTs was constant from the 3 ($M = .569$, $SD = .061$) to the 10-year follow-up ($M = .580$, $SD = .062$), but moved downward at 15-years ($M = .556$, $SD = .050$), with the difference between the 10th and the 15th year mean AUCs reaching significance, $t(26) = 3.20$, $p < .05$.

For victim-involved non-sexual offenses the temporal pattern for the BOs was the same as for the sexual charges. There was a non-significant dip in predictability for the BOs from the 3 ($M = .608$, $SD = .061$) to the 10-year period ($M = .604$, $SD = .061$), but the AUCs recovered at 15-years ($M = .614$, $SD = .064$), with the 15-year gate posting a slight superiority to the 10-year

AUCs, $t(26) = 2.87$, $p < .01$. In contrast, the BTs showed consistent increase over the three time blocks, with all three comparisons reaching significance, $t(26) = 5.99$, $p < .001$ for the 3 and 10-year comparison, $t(26) = 9.07$, $p < .001$ for the 3 and 15-year comparison, and $t(26) = 5.38$, $p < .001$ for the 10 and 15-year comparison (see Table 6).

For the BOs, the temporal pattern for victimless crimes was similar to that found for the BOs for the other crime types, except that the initial dip from 3-years to 10-years was significant for this crime type, $t(26) = 6.18$, $p < .01$. Between the 10th and 15th years, the predictability returned to the 3-year level, $t(26) = 5.74$, $p < .001$. For the BTs, in contrast, AUCs gradually increased, with the 3-year to 10-year, $t(26) = 2.62$, $p < .05$, and 3-year to 15-year comparisons, $t(26) = 4.28$, $p < .001$, reaching significance.

Child Molesters and Rapists

Child Molester and Rapist Comparisons in Recidivism

Comparative Kaplan-Meier survival analyses between rapist and child molester subgroups for the entire sample revealed that the former group recidivated at a faster and higher rate than the latter group for both non-sexual victim-involved and victimless charges, $LR(1) = 12.41$, $p < .001$, and $LR(1) = 9.21$, $p < .01$, respectively (see Figures 8 and 9). For serious sexual charges no difference in rate or level of recidivism emerged, $LR(1) = .16$, *ns* (see Figure 7). Whereas the recidivism rates for serious sexual charges for the rapists using the 3, 10, and 15-year gating criteria were 12% (18 of 154), 20% (21 of 104), and 20% (11 of 56), respectively, the *projected* estimates from the survival curves for the same time periods were 12% (23 of 191), 19% (36 of 191), and 20% (38 of 191). Similarly, the recidivism rates for serious sexual charges for the child molesters using the 3, 10, and 15-year gating criteria were 12% (24 of 197), 20%

(31 of 152), and 20% (18 of 91), with the projected survival analyses estimates for the same time gates being, 12% (30 of 251), 18% (45 of 251), and 24% (60 of 251).

Rapist and Child Molester Comparisons in Predictability

The comparisons on the overall prediction of serious sexual charges for the rapist and child molester subgroups were informative. Using the *t-test* strategy described earlier, we compared the mean sexual charge AUCs of the rapists and child molester groups across the 27 predictors for the 3, 10, and 15-years. Whereas the average AUC was significantly greater for the rapists than for the child molesters at the 3-year gate, $t(26) = 3.88, p < .001$, no differences were evident at the 10 and 15-year gates, $t(26) = .33$ and $.18$, respectively (see Table 6).

Non-sexual charges yielded a somewhat different pattern. For victim-involved non-sexual crimes, rapists were better predicted than child molesters at every temporal gate, $t(26) = 10.31, p < .001$, $t(26) = 4.30, p < .001$, and $t(26) = 2.81, p < .01$, respectively for the 3, 10, and 15-years follow-up gates. A similar pattern emerged for the victimless, non-sexual charges at the 3-year, $t(26) = 2.63, p < .05$, and 10-year gates, $t(26) = 3.33, p < .01$, but no differences were found at the 15-year comparison, $t(26) = -.572$.

Rapist and Child Molester Comparisons of Predictability across Outcome Time

Rapists and child molesters showed a different cross-temporal pattern of predictability on the predictive scales. Predictability significantly *increased* for the child molesters from the 3 to the 10-year gate, $t(26) = 4.26, p < .001$, and the level of predictability did not change from the 10-year to the 15-year gate, $t(26) = -.307, ns$. In contrast, predictability for rapists *decreased* the 3 to the 10-year gate, $t(26) = -2.32, p < .05$, and the level of predictability remained constant from the 10-year to the 15-year gate, $t(26) = -.492, ns$. Thus, the general trend for sexual

charges is for predictability for child molesters to increase over time and for rapists to decrease over time.

A different pattern was found for the non-sexual crimes. For victim-involved charges rapists' level of predictability remained constant across the three temporal gates, $t(26) = 1.19, 1.94,$ and $1.64,$ for the 3 to 10-year, 3 to 15-year, and 10 to 15-year comparisons, respectively. Child molesters' AUCs significantly increased across time, with all three temporal comparisons reaching significance, $t(26) = 7.65, 7.65,$ and $4.02, p < .001,$ for the 3 to 10-year, 3 to 15-year, and 10 to 15-year gates, respectively. A somewhat similar pattern emerged for the victimless non-sexual charges. Again rapists' AUC means remained constant across the three temporal gates, $t(26) = -.91, .55,$ and $1.57,$ all *ns*, for the 3 to 10-year, 3 to 15-year, and 10 to 15-year gates comparisons, respectively. Child molesters showed no change in predictability from the 3 to 10-year gates, $t(26) = .61, ns,$ but their AUCs increased significantly from 10 to 15-years, $t(26) = 8.00, p < .001.$ Thus, the general trend for non-sexual crimes predictability is for rapists to remain constant and for child molesters to increase.

Differential Predictability of Specific Crime Types

We compared the serious sexual charge mean AUCs to both kinds of non-sexual crimes, victim-involved and victimless, to determine the comparative predictive potency of the various scales listed in Table 6 for the entire sample, and for the rapists and child molesters separately at each of the three temporal gates assessed, 3, 10, and 15-years. The results of these comparisons are presented in Table 7. Basically, the measures predicted subsequent serious sexual charges better than victim-involved charges for all three groups at 3 years, and predicted serious sexual charges better than victimless charges for the entire sample and rapists, but not for the child molesters at this 3-year gate. No differences in predictive potency between serious sexual and

victim-involved charges emerged for any group at the 10 and 15-year gates. For the comparisons between serious sexual and victimless offenses at the 10 and 15-year gates, serious sexual charges yielded superior prediction only in the 10-year gate for the entire sample and for child molesters. These differences were all small effect sizes.

We selected only those measures that had been specifically fashioned to predict sexual recidivism and recalculated the same comparisons presented in Table 7. The results of these analyses are presented in Table 8. In general using this selective subset increased both the effect sizes of the comparisons presented in Table 7 and the frequency with which comparisons reached significance. In these comparisons the measures predicted serious sexual charges better than both kinds of non-sexual charges for all time gates for the entire sample. For rapists the sexual-recidivism-specific actuarials yielded superior AUCs for serious sexual charges as compared to both non-sexual charges at the 3-year gate. At the 10-year gate their prediction of the serious sexual charges was only superior to the victimless offenses, and no predictive differences emerged and 15-years. For child molesters the sexual-recidivism-specific actuarials yielded superior AUCs for serious sexual charges in all comparisons *except* for victimless charges at the 3-year gate.

Next, we selected those measures that had been specifically fashioned to predict or identify general criminality and recalculated the same comparisons presented in Table 7. The results of these analyses are presented in Table 9. For the entire sample these criminality-related measures predicted both non-sexual charges better only at the 15-year gate. For rapists they predicted only victim-involved charges better at the 10 and 15-year gates. For child molesters they predicted only victimless charges better at the 15-year gate.

2. Meaningful Differences Among Modern Risk Assessment Instruments in their Level of Predictive Accuracy

Table 5 presents the confidence limits for the AUC values for serious sexual charges for 3, 10, and 15 years for the entire MTC sample. It can be readily observed that the confidence limits for the AUCs overlap substantially for most of the scales, indicating that most of the variation in the differences in predictive accuracy can be attributed to chance. The confidence intervals for the AUCs for the two non-sexual charges yielded comparable results. Nonetheless, we thought that it was important to identify the patterns of differences and similarities that might exist. Consequently, for each charge type at each outcome gate we calculated the significance of the difference between the AUC values. Because the AUCs were calculated on the same samples, we used the nonparametric method proposed by DeLong, DeLong, and Clarke-Pearson (1988) for comparison of correlated samples. This procedure uses the Mann-Whitney U statistic proposed by Hanley and McNeil (1983) to derive the AUC and its *SE* for each of the correlated curves. We then calculated the covariance and correlation matrices of the curves and used a method of structural components analysis to contrast the curves, deriving a chi-square statistic with degrees of freedom equal to the number of tests being examined minus one. We calculated two-tailed *p* values, comparing curves at the .05, .01, and .001 levels. The nine tables with the results of comparing each charge type at each outcome gate are available from the authors. We will simply describe here the major patterns of differences that emerged.

Comparison of the AUCs for Serious Sexual Charges

For all crime types we will focus on the 3 and 10-year gate results, because by the 15-year gate the number of offenders (maximum BTs = 63 and BOs = 149; 387 missing) and therefore power had decreased sufficiently that few differences reached significance, even those

with large effect sizes. For both the 3 and 10-year comparisons almost all the differences found for predicting sexual charges could be accounted for by five measures that were consistently *below* the other measures. No differences emerged among the best predictors. Table 10 presents these five measures and the number of comparisons in which they were significantly *below* other measures at the .05, .01, and .001 levels for the 3 and 10-year gates.

The two measures with the poorest relative predictive accuracy were two A-SOAP measures-- A-SOAP Intervention, Part 3 and A-SOAP Dynamic Total. The Dynamic Total score included the A-SOAP Intervention measure (Part 3) and the A-SOAP Community Stability score (Part 4). The Community Stability score was not rated for many offenders, because it could only be scored when information on the adaptation of the offender in the community was available, and such data were found in the files of only a small number of offenders. Not only were these scores not significantly related to outcome at any time gate (see Table 6), but, as can be seen in Table 10, for both the 3 and 10-year gates both scores were significantly less predictive than many of the other measures, with a substantial number of these comparisons exceeding .001 significance (e.g., for both measures at the 10-year gate they evidenced lower AUCs than 19 other measures, $p < .001$ [see Table 10]).

Relative to other measures the RM2000 Violence Index also did not fare well in predicting serious sexual charges. Not only was it not significantly accurate at the 3, 10, and 15-year gates in predicting sexual charges (see Table 6), but at the 10-year gate it has significantly lower AUCs than 18 of the other measures, and 7 of these comparisons reached .001 significance (see Table 10).

Although the differences of the SRA Need Assessment with other measures only reached significance when compared with the poorer predictors discussed above, nonetheless it is

noteworthy that this is the only measure that achieved AUCs greater than .700 at each of the time gates assessed, reaching an AUC of .756 ($p < .001$, $CI = .669 - .843$) at the 15-year gate. It is also noteworthy that the three conceptual mechanical risk assessment scales, the PCL-R, ASOAP-II, and SVR-20, performed only moderately well in predicting subsequent serious sexual recidivism, yielding AUCs that ranged from .641 to .684 across the three time gates for their total scores. Their subscales, especially the PCL-R Factor 1 and the A-SOAP Intervention scale fared somewhat more poorly, often falling significantly below other predictors. Although the total scores for these three conceptually driven scales did not differ significantly from the mechanical actuarials, they were consistently a little lower than these instruments in predicting sexual recidivism.

Comparison of the AUCs for Victim-Involved Charges

For both the 3 and 10-year comparisons almost all the differences found between the measures for predicting victim-involved charges could be accounted for by seven measures for the 3- and 10-year gates. As can be seen in Table 11, three measures—the RM2000 Violence and Combined Risks and the Mn-SOST-R yielded significantly higher AUCs than a number of measures, and the A-SOAP Sexual Part 1, the SRA sexualization score, the Doren, and the RRASOR afforded lower AUCs. At both time gates the RM2000 Violence and Combined Risks yielded *higher* AUCs than a considerable number of other measures. Clearly, because the latter measure contains the former, and the RM2000 Violence Risk yielded more significant comparisons, the Violence Risk accounts for the predictive potency in the combined score. The Violence Risk was significantly higher than 16 other measures at the 3-year gate and 12 measures at the 10-year gate. Slightly below these measures was the Mn-SOST-R, which

predicted victim-involved charges better than 8 and 7 measures at the 3 and 10-year gates, respectively.

Of the four measures that were significantly lower than other predictors, the two focusing on sexualization (A-SOAP Sexual Part 1 and the SRA sexualization) were the worst measures at predicting non-sexual victim-involved crimes. At the 10-year gate the former was inferior to 19 other measures, and the latter was significantly lower than 20 other measures. The Doren, which uses a decision algorithm that combines the Static 99 and the RRASOR, was significantly inferior to 4 and 8 other measures, respectively, at the 3 and 10-year gates (see Table 11).

Comparison of the AUCs for Victimless Charges

The differences in predicting non-sexual victimless charges added substantially to the list of poor predictors generated for victim-involved charges and yielded larger, as well as more extensive, measurement differences. We selected all those measures that were significantly lower than at least one other measure at $p < .001$ at both the 3 and 10-year gates. Table 12 presents the number of comparisons in which these chosen measures were significantly lower than other measures at three significance levels for the 3 and 10-year gates. Represented among these poor predictors of victimless charges were the measures of sexualization, intervention, callousness-unemotionality (PCL-R Factor 1), early conduct disorder, and social-affective competence.

3. Can Cohesive, Meaningful Predictive Dimensions Be Identified

in the Extant Risk Assessment Instruments?

Factor Structure of the Actuarials

To explore the issue of what dimensions might underlie the extant actuarials, we calculated a principle component analysis with iterations on items selected from eight of the risk

assessment instruments (MnSOST-R, Static-99, Static-2002, RM2000/S, RM2000/V, SORAG, A-SOAP-II, and the SVR-20). Because these actuarials have been used interchangeably on rapist, child molesters, and mixed-victim-age offenders, the entire sample was analyzed together. On the basis of an examination of the Scree Plot, we chose the five-factor solution that accounted for 48.2% of the variance, and we rotated these factors to simple structure using Varimax. Table 13 presents the eigenvalues for each of the factors in the five-factor solution and the percent of variance accounted for by each factor.

Table 14 presents the rotated component matrix for the five-factor solution with items sorted by their loadings on the factors. For ease of examination only loadings $>.20$ are presented, and item loadings $>.50$ are color coded to match the factor on which they have their primary loading. Although several Factor 1 items have secondary loadings on Factor 2, and Factor 2 items similarly load on Factor 1, and the stranger items on Factor 4 have secondary loadings on Factor 2, but the expressive aggression items of Factor 4 items load secondarily on Factor 1, nonetheless, in general the solution is fairly clean and conforms reasonably well to simple structure (i.e., each item loads high predominately on one factor and low on the other factors).

Examination of items loading $>.50$ on the first factor suggest that it reflects persistent general criminality and resistance to rules, and can be appropriately labeled Criminal Persistence. It is notable that all the scales contribute items to this first factor, and the items reflect a range of different kinds of crime (violent and non-violent), the persistence of crime, juvenile as well as adult offending, and specifically, resistance to supervision.

Examination of the items on Factor 2 suggests that it reflects persistence and rate of sexual offending, and can be reasonably labeled Sexual Persistence. It includes items reflecting the various ways of coding amount of prior sexual offending, both the duration and rate of prior

sexual offending, and whether offending included non-contact sexual offenses. Non-contact sexual offending is a form of offending that is believed to be particularly persistent.

Items referring to youth and to never having been married loaded consistently and fairly exclusively on Factor 3. These items do not, however, represent a simple linear reporting of age, but rather reflect a compilation of the various cut-points used in coding age in different prediction instruments. Predictably, never having been married was related to being young, and consequently loaded on this factor. We gave this factor the descriptive name, Youth & Single.

Items that related both to having offended against strangers and to the extent of physical harm or abusive threatening behavior directed at victims loaded on Factor 4. As noted above, whereas the stranger items also loaded on the Sexual Persistence factor, the expressive aggression items on this factor loaded on Factor 1. The relation of violence generally and sexual violence specifically to the high externalizing behavior captured in Factor 1 is consistent with other research (e.g., Daversa, Sitnikov, & Knight, 2005; Poythress & Skeem, 2006). The covariation between stranger victims and violence in the offense that was sufficient to generate Factor 4 is most likely complexly determined, possibly involving a confound with the rapist-child molestation distinction. Rapists are both more likely to assault strangers and to exhibit significantly higher levels of aggression in their sexual assaults (Bard et al., 1987; Daversa et al., 2005), so in a sample like this one, where they are mixed, stranger and violence would tend to covary. It is also more common that contact sexual offenses against strangers involve more threats, physical coercion, or injury than offenses against known victims, because of the increased contextual demand for violence to achieve compliance in stranger assaults. A reasonable abbreviated label for this factor is Violent Stranger Assaults. Intermediate scores on this factor would likely result from either high physical violence used against a known victim or

an offense against a stranger where physical violence was absent. It is noteworthy that the VRAG victim injury low item refers to the lack of injury to victims and consequently loaded negatively on this factor.

Finally, Factor 5 refers to male victim choice. The combination of items indicates that offenders scoring highest on the factor had male child victims and no female victims. Those with an adult male victim or a mixture of male and female victims would likely be the next highest scoring group. Those with only female victims would have scored lowest on this factor. This factor was labeled Male Victim Choice.

Predictive Potency of the Five Factors

Scale scores for each factor were created by standardizing all items that loaded $>.50$ and calculating their mean. Only offenders with complete data were included. Consequently, the N for the 5-year follow-up analyses was 358, whereas 10-year follow-up had 294 participants. Table 15 shows the resultant predictive accuracy for each of these factors and the confidence intervals around the estimated AUCs.

Only the Young & Single factor was not significantly related to sexual recidivism, yielding an AUC that was essentially at chance level ($AUC = .494$, $CI = .420-.569$). All the other factors afforded statistically significant predictive accuracy with their lower confidence limits exceeding chance. Whereas Sexual Persistence and Male Victim Choice both had AUCs in the mid .60s, Criminal Persistence and Violent Sexual Assaults had AUCs just below .60. There was, however, substantial overlap in their confidence limits of these four factors.

Predictive Potency of a Composite of the Five Factors

The Varimax rotation produced factors that minimized the intercorrelations among factors, thereby maximizing their potential for enhanced predictive potency when combined. A

summative composite, which was created by weighting each factor equally, yielded an AUC of .702 ($CI = .648 - .772$) for a 5-year follow-up and an AUC of .690 ($CI = .621 - .759$) for the 10-year follow-up. This composite of equally weighted factors showed predictive accuracy at least as good as that of any of the prediction instruments for which results had been presented in Table 5.

Based on *a priori* considerations (and before examining their actual predictive value in this sample), we had postulated that the Sexual Persistence factor should be double weighted relative to the other factors. A composite of factors weighted in this way achieved slightly greater predictive accuracy. For the 5-year follow-up it was .710 ($CI = .648 - .772$), and for the 10-year follow-up it was .708 ($CI = .641 - .774$).

Percent of Variance of the Actuarials Accounted for by the 5-Factor Solution

Table 16 shows the correlations between the five static historical factors and scores on each of the established mechanical actuarial instruments. Clearly, the Sexual Persistence factor was the most highly correlated of the factors with these instruments (Average $r = .58$), closely followed by the Criminal Persistence factor (Average $r = .48$). To assess the total variance of the actuarials accounted for by the factors, six ordinary least squares regression equations were calculated using sequentially the six actuarials as dependent measures and, the five derived factors as independent measures in all the analyses. Table 17 presents the percentage of variance accounted for in each actuarial by the five factors (R^2). Most of the variance of these scales was accounted for by the five static historical factors.

Residual Predictive Variance in the Actuarials over and above the 5 Factors

Twelve logistic regression equations predicting sexual recidivism (six each for each of the two time gates) were calculated to assess whether any of the mechanical actuarials

established specifically for predicting sexual recidivism (MnSOST-R, Static-99, Static-2002, RRASOR, RM2000/S, and SORAG) contributed unique predictive variance over and above the five factors. The five factors were entered in the first step of the logistic regression and the change in the log-likelihood ratio chi square was assessed when each mechanical actuarial was entered in the second step. None of the second steps even approached significance, indicating that no residual predictive variance was accounted for by any of these mechanical actuarials.

Similar logistic regressions were calculated using the three conceptual mechanical instruments (PCL-R total, A-SOAP II total, and SVR-20). There was a non-significant trend ($\chi^2(1) = 3.11, p < .10$) for the PCL-R total score to add significantly to the prediction of sexual recidivism at the 5-year gate, which was *not* repeated at the 10-year gate. Moreover, SVR-20 significantly contributed to prediction of sexual recidivism at the 10-year gate ($\chi^2(1) = 3.96, p < .05$), though it did not add significantly to prediction at the 5-year follow-up. The A-SOAP II did not add significantly to the predictive accuracy of the five static historical dimensions.

4. Predictive Contribution of the SRA Need Framework

Earlier, in the section comparing the predictive potency of the various actuarial measures, it was noted that the SRA Need Assessment fared extremely well in comparison to other actuarials. As indicated, although it did not differ significantly from many of the better performing actuarials, nonetheless, SRA Need Assessment yielded the highest AUCs for the entire sample for predicting sexual recidivism. It is important to assess whether the SRA Need Assessment accounts for independent variance in predicting recidivism. To determine its independent contribution we calculated logistic regressions predicting sexual recidivism at 5 and 10-years. The SRA Need Assessment was entered in the second step after the five factors had been entered in the first step. In both analyses SRA Need significantly contributed independently

to the prediction of sexual recidivism, $\chi^2(1) = 6.06, p < .025$ and $\chi^2(1) = 9.43, p < .005$, for the 5 and 10-year analyses, respectively.

To get a better idea of how Need combined with the static historical factors, we calculated two further logistic regression equations. Both of these had two predictors: the SRA Need Assessment score and a combination score of the static historical factors formed by double-weighting Sexual Persistence and weighting the other factors equally. The correlation between these two variables was significant, $r(456) = .505, p < .001$, indicating the possibility that shared variance might also contribute to predicting outcome. The regression weights for each predictor shown in Table 18 reflect the independent contribution of each scale to predicting sexual recidivism at each outcome gate. To aid interpretation of the regression weights and the E(B) statistics, each factor was converted to a z score prior to fitting the equation. It is apparent from the betas presented in Table 18 that each predictor made an independent contribution to the prediction of sexual recidivism at each time gate. Whereas the static historical factor did slightly better in predicting sexual recidivism at 5 years, the Need factor was slightly superior at predicting recidivism at 10 years.

Predictive Accuracy of Combining the Composite Historical Score and the Need Factor

The two predictors from the previous equation were converted to z scores and summed to give a combined predictor that equally weighted static historical composite and the Need scale. Table 19 presents the AUCs predicting serious sexual recidivism for the 5 and 10-year gates. Adding the SRA Need Assessment score appears to lead to materially better prediction than predicting from the static historical variables alone.

5. Relation between Age and Sexual Recidivism

Recently a debate has emerged about what role the age of the offender should play in applying actuarials. Some have argued that it should be considered an important moderator of recidivism, because older offenders have been found less likely to recidivate (Barbaree, 2003; Prentky & Lee, in press). In the present study the third factor from the static historical variables (Young & Single) had no useful predictive value. Similarly, when logistic regression equations predicting sexual recidivism from age on discharge were fitted the E(B) coefficients were 0.997 and 0.998 for the 5 and 10-year sexual recidivism outcomes, respectively. Not only are these coefficients not significantly different from 1.0, but also their magnitude indicates that aging even by 20 years produced only a small, non-significant reduction in sexual recidivism. When we divided age at discharge into decades (analogous to the division used by Prentky and Lee [in press]), reversed the ordinal direction (because a younger age is suppose to predict higher probability of recidivism), and used this scale to predict serious sexual recidivism at the 3 and 10-year gates, the *AUCs* were .527 (*CI* = .414 - .640) and .511 (*CI* = .405 - .617), respectively. Because these results were different from what other studies have found, we examined the data for a more complex effect.

First, we coded age at index offense according to the SORAG scoring schemes, so that higher scores meant a younger age. This scale could be described as the youthful offense age at index offense (Youth at Index Offense). We regressed this against serious sexual recidivism at the 5 and 10-year time gates, but it yielded no statistically significant results. Younger age at index offense was not related to recidivism.

Next, age on discharge was added to the equation (coded in years). In these equations Youth at Index Offense yielded a positive regression coefficient that was statistically significant

($b = .176, p = .045$) for sexual recidivism at 5-year outcome, and nearly so ($b = .172, p = .069$) for 10-year outcome. These results indicated that when age on discharge was controlled, those who were younger at their index offense were more likely to recidivate sexually. In contrast, neither of the coefficients for age on discharge was significant and both were also positive, indicating that (if anything) sexual recidivism rates increased as offenders' age increased beyond what it had been at their index offense.

In a third step, the quadratic effect of age on discharge was added. This led to significant coefficients for all three variables in the equations for sexual recidivism at both the 5- and 10-year. Table 20 presents the relevant coefficients. These results suggest two distinct effects. Sexual recidivism rates were higher for those who were younger at their index offense, but overlaid on this was a quadratic effect of age on discharge, such that (holding age on index constant), older age on discharge was initially associated with raised sexual recidivism, but beyond a certain point, this trend reversed.

To clarify the nature of these results Table 21 explicates the results produced by banding the two aspects of age and plotting serious sexual recidivism as a function of the different combinations of age on index offense and age on discharge. Caution in interpreting 60+ age-at-discharge group is advised because the n 's here were very small. What is striking is that for every comparison where there were sufficient data, except the 60+ group, older age at discharge was associated with higher sexual recidivism, once age at index offense was controlled.

6. Determining Differential Predictors for Rapists and Child Molesters

In general comparisons on extant actuarials have not typically yielded significant differences in predicting recidivism between rapists and child molesters. In light of our analysis of the core components in these actuarials, in which Criminal Persistence and Sexual Persistence accounted for the majority of the shared variance, this is not surprising. There is no reason *a priori* to hypothesize that these two factors should predict differently for the two samples. The only factor that would be hypothesized to predict differently for the two samples would be Male Victim Chosen, which correlated substantially only with the RRASOR. Nonetheless, in the present study we compared rapists and child molesters on all of the predictive measures listed in Table 5. Because the comparisons were calculated between independent samples, we used the *z* statistic (Vida, 2006), which was obtained by dividing the difference between the AUCs by the standard error of the difference.

For serious sexual recidivism only three group comparisons reached or approached significance and these were all at the 3-year outcome gate. The AUCs of child molesters were lower for the A-SOAP II Intervention ($AUC = .431$ [$CI = .299 - .568$] and $AUC = .668$ [$CI = .530 - .806$], for child molesters and rapists, respectively, $z = 2.36$, $p < .025$) and Dynamic total subscales ($AUC = .438$ [$CI = .307 - .568$] and $AUC = .665$ [$CI = .528 - .803$], for child molesters and rapists, respectively, $z = 2.31$, $p < .025$), and higher for the SRA sexualization scale ($AUC = .730$ [$CI = .615 - .846$] and $AUC = .549$ [$CI = .407 - .688$], for child molesters and rapists, respectively, $z = 1.94$, $p = .052$). For victim-involved, non-sexual charges only two group comparisons reached significance and both of these involved the SRA Need Assessment score. For both the 3 ($z = 2.28$, $p < .025$) and 10-year ($z = 2.33$, $p < .025$) gates the SRA Need Assessment score AUCs for rapists ($.715$ [$CI = .625 - .804$] and $.757$ [$CI = .663 - .852$],

respectively for the 3 and 10-year gates) were higher than those for child molesters (.538 [$CI = .414 - .662$] and .569 [$CI = .441 - .697$], respectively). For victimless offenses only one comparison reached significance. At the 3-year follow-up the RM2000 Violent Risk score predicted victimless charges better for the rapists (.719, $CI = .647 - .791$) than for child molesters (.648, $CI = .556 - .739$), $z = 2.26$, $p < .025$). Because of the large number of comparisons made between these groups, these differences could certainly have arisen simply by chance.

Consistent with the low frequency of rapist-child molester differences in AUCs for predictive measures, in logistic regressions predicting the 5 and 10-year serious sexual outcome, when a dichotomous victim-age-choice variable was stepped in either after the composite five factor score or the SRA Need Assessment score, no significant increases in predictability emerged. Moreover, stepping in the interaction of this victim-age-choice dichotomy with either the composite five factor score or the SRA Need Assessment score likewise yielded no increases in predictive potency. As we saw earlier, the survival curves for rapists and child molesters for serious sexual recidivism were equivalent, so one would not expect such a dichotomous variable to predict outcome. Moreover, the extant actuarials do not target domains that should be differentially related to outcome in these two offender groups. Consequently, these analyses do not satisfactorily address the question of the differential predictability of rape and child molestation.

To address this issue more thoroughly we first used sets of predictors that were hypothesized to be differentially related to recidivism within each age-preference group to predict outcome at multiple outcome time gates using DFAs. These analyses allowed us to calculate the linear composite at each time gate that best discriminated those who recidivated from those who did not for each age-preference group, and to identify the variables that loaded

highest on these composites. We then used Cox hierarchical regression analyses to determine whether the critical predictors we identified were able to improve prediction over the actuarial that had been determined in the prior analyses to be the best predictor of sexual recidivism for both rapists and child molesters, the SRA Need Assessment score.

Tables 22 and 23 present the results of the DFAs predicting serious sexual charges for child molesters and rapists calculated at each of five time gates. The predictors were derived from the MTC Coding dictionary and the MTC Child Molester (MTC:CM3) and Rapist (MTC:R3) typology ratings and are described in Appendix 3. Justification for the selection of predictors can be found in prior publications (Bard et al., 1987; Knight, 1999; Knight & Guay, 2006; Knight, Rosenberg, & Schneider, 1985; Knight & Sims-Knight, 2003, 2004; Prentky & Knight, 1991; Prentky, Knight, & Lee, 1997). For child molesters paraphilias, fixation, social isolation, adult sexual crimes against children, presence of at least one male victim, and impulsivity were selected as initial independent variables (see Appendix 3). For rapists pervasive anger, offense planning, juvenile unsocialized behavior, adult unsocialized behavior, Sadism, and adult sexual crimes against women were chosen as independent variables (see Appendix 3). Because many of these variables were derived from the dimensional judgments in the MTC typologies, the variables were only rated when an offender was categorized within the individual typology. Rapists were not typed as child molesters, and vice versa. Consequently, we were not able to do comparisons on these variables across rapists and child molesters.

All DFAs were stepwise with both forward selection and backward removal. Consistent with the AUC analyses presented in Section 1 of the Results, for child molesters (Table 22) canonical correlations increased at longer time gates with the highest canonical correlations achieved at 7 and 10-year follow-up gates. Five of the six variables contributed to defining the

discriminant functions at various times. Only Male Victim failed to contribute to defining the DFAs at any of the gates. The DFA at the 3-year time gate was defined by variables measuring sexualization (paraphilias) and sexual preference (high fixation). In the middle period impulsivity and the frequency of adult sexual crimes against children entered the picture. Finally, social isolation at the 10 and 15-year gates emerged as the major contributor to sexual recidivism. These analyses suggest both that the prediction of outcome increases over time for child molesters, and that variables optimal for predicting early recidivism might not contribute to later prediction.

For the rapists (Table 23) a very different picture emerged. Consistent with the analyses of the AUCs, prediction of sexual recidivism appeared to decline over time, with the 3-year gate yielding the highest canonical correlation. The same two variables--pervasive anger and offense planning--were the only variables operative in defining the DFAs. Pervasive anger was the stronger predictor at the 3 and 5-year gates, but it dropped out after the 5-year gate, leaving offense planning to carry the defining of the DFAs at the 7 and 10-year follow-up periods. At the 15-year gate only offense planning approached significance.

To determine whether the five significant independent variables for child molesters and the two significant independent variables for rapists contributed predictive variance that was independent of the extant actuarials, two hierarchical Cox regression analyses were calculated. In each the SRA Need Assessment score, which consistently showed the highest AUCs of all the actuarials for both the child molesters and rapists at each time gate, was entered in the first step and the group specific predictors were entered in the second step. For the child molesters, as expected, in the first stage the SRA Need Assessment score significantly predicted the serious sexual recidivism hazard rate, $\chi^2(1) = 7.48, p < .01$. More importantly, however, the additional

independent variables entered at the second step accounted for considerable additional independent variance, $\chi^2(5) = 31.64, p < .0001$. The final hazard ratios for the six predictors, SRA Need Assessment, paraphilias, social isolation, fixation, impulsivity, and adult sex crimes, were 1.02, 1.51, 2.09, 7.57, 2.42, and 1.12, respectively.

For the rapists at the first stage, the SRA Need Assessment score significantly predicted the serious sexual recidivism hazard rate, $\chi^2(1) = 13.96, p < .001$. In the second stage, however, entering pervasive anger and offense planning did not account for additional predictive variance, $\chi^2(2) = 4.57, p = .102$. It was obvious in doing the file ratings that the data on both pervasive anger and offense planning was insufficient in the BO files, whereas information on the BTs was more complete for these two variables. Consequently, we redid the regression analysis on only the BTs. Once again in the first stage the SRA Need Assessment score significantly predicted the sexual recidivism hazard rate, $\chi^2(1) = 8.71, p < .005$. In contrast to the analysis of the entire sample, the addition of pervasive anger and offense planning did contribute independent variance to prediction in this analysis, $\chi^2(1) = 11.16, p < .005$. The hazard ratios for the three predictors, SRA Need Assessment, pervasive anger, and offense planning, were 1.10, 1.32, and 1.34, respectively. These results suggest that the two variables might account for independent predictive variance, but only when sufficient data for their evaluation are present.

Focus on issues of specialization and generalization might provide another avenue for enhancing prediction among child molesters. Harris, Knight, Smallbone, & Dennison (2006) have examined the criminal histories of the offenders in this sample, applying to these histories various definitions of specialization, the tendency to repeat the same offence type on successive arrests. Using a specialization threshold definition of at least 50% of their offenses being sexual, 31.5% of child molesters, but only 9.3% of rapists would be considered specialists. Harris is

currently working on creating a more sensitive specialization criterion that takes into account developmental criminal history and specialization over the lifespan. Clearly, one would hypothesize that different variables are likely to predict for specialists and generalists.

DISCUSSION

Overview of Findings

We will organize the discussion in the same way that we presented the results.

1. Reliability and Predictive Potency of Actuarials Applied to the MTC Sample

Although the record sources accessed in the present project were written at a time that preceded the recent research attention to prediction for sexual offenders and consequently did not focus on and detail those variables that have now achieved greater salience among practitioners working in the area, and although these records sometimes provided vague or incomplete prior criminal histories, which are so critical to rating many actuarials, nonetheless the interrater reliabilities achieved by the study's raters and the overall level of predictive accuracy obtained were comparable to other independent evaluations of actuarials (e.g., Langton, 2003). This accuracy was also remarkable in light of the difficulty collecting accurate and comprehensive recidivism data in the USA as compared to the UK, Canada, or Sweden where there are high quality national conviction databases available to researchers. This follow-up problem was likely addressed to some degree because of the strategy in the present study of accessing and integrating multiple outcome sources. Finally, because all of the offenders in the sample had been referred for a legal determination of whether they should be committed to MTC offenders who were obviously at low risk were likely screened out, and their advantage to enhancing prediction was lost. As noted earlier, the average risk level of the population was higher than is normally found in a prison sex offender population.

Overall, modern instruments designed for use with sexual offenders showed a moderate level of accuracy predicting sexual, victim-involved, and victimless charges at various time gates. The large number of predictive instruments and the relatively long outcome period allowed us to examine systematically several issues about the temporal stability of predictors, the differential predictability of subsamples, and the comparative prediction of different crime types. We will summarize in turn our results in each of these domains.

Cross-temporal Stability of Prediction

Over the entire sample the average prediction of serious sexual recidivism appeared relatively temporally stable. In contrast, the accuracy of the prediction of victim-involved offenses increased at each time gate examined, and the accuracy of victimless offenses was stable through 10 years, but increased at the 15-gate.

Comparisons between the BO and BT Samples

The present BO sample (those declared not sexually dangerous and not committed) comprised two subsamples, one matched and one randomly selected subsample of all BOs evaluated at MTC. The BTs had both a more rapid and higher rate of recidivism than both the entire BO and the randomly selected BO samples. Estimated recidivism rates from the survival analyses indicated 32% recidivism for the BTs at 15 years and 12.5% for the randomly selected BOs. For victim-involved offenses BTs had more rapid and higher recidivism rates only when compared to the random BOs, but not when compared to the entire BO sample. BTs differed from neither the random BO sample nor the entire sample in their subsequent rate of victimless offending.

The issue of the actual base rate of the sample for whom one is attempting to predict is an important one and has been a controversial part of the debate about the accuracy of civil

commitment judgments relative to base rates (Janus & Meehl, 1997; Doren & Epperson, 2001). The incremental validity of an actuarial predictor over base rates is directly related to the base rate of the phenomenon, such that base rates closer to 50% yield the higher incremental validity for particular actuarial instrument of known validity (Meehl & Rosen, 1955). Citing low base rates of Janus and Meehl (1997) have criticized current judicial civil commitment decisions. Doren and Epperson (2001) extrapolated higher recidivism base rates and criticized the accuracy of Janus et al.'s analyses. The best estimate from the present study to address this issue is the combined projected 15-year rate of recidivism (21.5%), where we have the most adequate sample for the longest duration. The combined group is the most appropriate sample for the civil commitment issue, because it contains adequate representation of offenders who were considered for civil commitment. This recidivism rate is more in keeping with those used by Janus and Meehl (1997) rather than those extrapolated by Doren and Epperson, (2001). Consequently, according to our outcome estimates Doren and Epperson's (2001) reanalysis has not adequately countered Janus and Meehl's (1997) original critique of current judicial civil commitment decisions

In general, and contrary to what might be expected from certain perspectives, BOs serious sexual charges were on average more predictable than BTs, with the differences in average AUCs reaching significance at two time gates and approaching significance on the third. From a Bayesian perspective a smaller probability recidivism target should be harder to predict than a larger one (Barbaree, 1997; Meehl & Rosen, 1955). Consequently, one might hypothesize that the higher recidivism rate of the BTs would allow greater accuracy of prediction. This was not the case. Interestingly, although across the board the means for the actuarials were significantly higher for the BTs than the BOs, the variances of these measures were comparable

across the groups, suggesting the possibility that it might be more difficult to predict recidivism among higher risk individuals with the current actuarials. The evidence for slightly higher predictability for the entire sample than for either subsample argues as well that greater heterogeneity of risk might enhance predictive accuracy.

For the non-sexual crimes the BOs tended also to show higher predictability when differences emerged, but the results were not as consistent as they were for sexual crimes. The general conclusion across crime types was that the entire sample yielded the highest AUCs, followed closely by the BOs.

Different temporal patterns in predictability emerged for the BOs and BTs. For all crime types the BOs were fairly consistent across time, showing a slight dip in predictability between the 3 and 10-years gates and recovery of early predictive levels at 15 years. In contrast, the BTs showed a slight dip in their predictability between the 10 and 15-year periods for sexual recidivism, but rather consistent increases in predictability over the 15-year follow-up for the two non-sexual crimes.

Comparisons between Child Molesters and Rapists

No differences emerged for the overall speed and frequency of sexual recidivism for the child molesters and rapists, with survival analysis estimated rates at 15 years being 24% and 20%, respectively for child molesters and rapists in the entire sample. In contrast, consistent with the hypothesis that rapists are criminologically more generalists than child molesters, rapists had significantly higher and faster rates of recidivism for both non-sexual crimes.

Differences in the temporal pattern of prediction emerged between the two groups. Rapists showed better initial predictability for sexual recidivism than child molesters, but this superiority dissipated over time. The cross-temporal analysis within each group showed

predictability decreasing for the rapists between the 3rd and 10th year and increasing for the child molesters during this same period. Both remained consistent from the 10th to the 15th year. For most comparisons on non-sexual charges, rapists were better predicted across the follow-up than child molesters, with the cross-temporal predictions remaining constant for rapists and increasing for child molesters.

Comparisons of Different Crime Types

Although the results here appeared complex, the basic trend is that measures fashioned to predict sexual recidivism predict such better than they predict non-sexual crimes, and those created to predict criminality or psychopathy are slightly better at predicting subsequent non-sexual than sexual charges.

2. Meaningful Differences Among Modern Risk Assessment Instruments

in their Level of Predictive Accuracy

The 3-year gate AUC coefficients for actuarials that were established empirically to predict sexual recidivism were approximately in the .670 to .700 range. They thus had relatively similar levels of predictive accuracy. These rates compare favorably with other follow-up studies (e.g., Craig et al., 2006; Harris et al., 2003; Langton, 2003). Across the extant studies no measure has arisen that has been consistently superior to others, with all measures varying substantially across studies. The SRA Need Assessment scale emerged as the highest predictor of sexual recidivism across the outcome periods, but it was significantly better than only the poorest measures.

Conceptual-mechanical (i.e., structured clinical guidelines) risk assessment instruments also had moderate though slightly lower levels of predictive accuracy. AUC coefficients were generally between .640 and .660 for the entire sample at the initial time gate assessed.

Detailed comparisons identified some actuarial subscales that fared more poorly than others in predicting sexual recidivism (e.g., the A-SOAP Dynamic scale and the RM2000 Violence Risk scale). Factor 1 of the PCL-R and the DSM-IV APD were also noteworthy as consistently poor predictors of sexual recidivism. Although the poor results for PCL-R Factor 1 may reflect the difficulty of coding this construct accurately from the kind of file data available in some of the archival records in this study, they are, nonetheless, consistent with the findings of other investigations of sexual offenders (Langton, 2003; Serin, Mailloux, & Malcolm, 2001), where Factor 2 predicted outcome and Factor 1 did not. Moreover, consistent with our results, in general criminal recidivism studies Factor 2 has fared better than Factor 1 in its predictive potency (Douglas, Vincent, & Edens, 2006; Hemphill, Hare, & Wong, 1998).

3. *Can Cohesive, Meaningful Predictive Dimensions Be Identified*

in the Extant Risk Assessment Instruments?

The factors generated from the items of the extant actuarials were broadly consistent with those found by Barbaree et al. (in press). They indicate that five interpretable dimensions—Criminal Persistence, Sexual Persistence, Young & Single, Violent Sexual Assault, and Male Victim Choice—can be used to summarize a large number of static historical items and that the resulting factor scores account for the predictive value of existing static risk assessment instruments. The first two factors consistently had the highest correlations with existing instruments, indicating the primary emphasis of the extant measures. Four of the five factors significantly predicted sexual recidivism. Consistent with the age analyses, Young & Single was the sole factor that was not related to sexual recidivism.

The results of this factor analysis should be tested using confirmatory factor analysis in other samples, but they suggest that future prediction instruments can be explicitly constructed to

measure the underlying dimensions. The results also open the possibility of integrating the results of existing instruments by identifying the underlying dimensions they assess and determining which operationalization of a factor yields the best reliability and validity.

4. *Predictive Contribution of the SRA Need Framework*

SRA Need Assessment yielded slightly higher AUCs than those of the best of the established risk assessment instruments. This is striking given the difficulty coders had in rating Need factors reliably from the MTC files. The SRA coding criteria require adjustments to improve reliability. Importantly, both static historical and Need variables made statistically significant contributions to prediction. An equally weighted combination of the two kinds of factors led to higher predictive accuracy, especially at the longer follow up (10 and 15 years).

5. *Relation between Age and Sexual Recidivism*

The present results suggest that age at discharge should not be introduced as a weighting factor for the actuarials. Strikingly, when age on index offense was controlled, increasing age on discharge was associated with increased rates of sexual recidivism., Those discharged after the age of 60, however, did have lower sexual recidivism rates, but the size of this sample was too small to make any firm recommendations. Nonetheless, the present results are consistent with adjusting expected recidivism rates down only for those discharged after the age of 60. The complexity of these results demand replication before too much weight is placed on them, but they are certainly inconsistent with the hypothesis that merely holding someone in prison will enable them to age out of risk.

6. *Determining Differential Predictors for Rapists and Child Molesters*

Consistent with the finding that the survival curves for rapists and child molesters in predicting sexual recidivism did not differ, the addition to regression models of a dichotomous

variable identifying the offender's status in these age-preference groups had no effect. In contrast, a strategy of focusing on subgroup specific measures that were either neglected or underutilized in extant actuarials and using these variables to predict subgroup-specific outcome yielded more promising results. Such measures were entered in DFAs to predict sexual recidivism at five time gates for rapists and child molesters separately. The earlier finding, when examining predictability over across numerous actuarial measures, that we might be able to predict more successfully the sexual recidivism of rapists in the short run and child molesters in the long run, was corroborated in these analyses. More importantly, five subtype-specific predictors for child molesters and two for rapists were identified. When these were entered into Cox Regressions to predict the hazard rate of sexual recidivism after the contribution of SRA Need Assessment had been removed, the five measures for child molesters accounted for significant additional predictive variance for the entire sample and the two measures for rapists accounted for significant additional predictive variance, but only for the BT sample, where they were more adequately measured. The measures included fixation on children, Paraphilias, male victim choice, Social Isolation, and Impulsivity for the child molesters, and pervasive anger and offense planning for the rapists.

Plans and Recommendations for Future Research

In general there are multiple avenues of research and analysis for improving the extant actuarials that the present study opens. Four of the findings in particular have important implications for future research.

First, the finding that common dimensions underlie static historical risk variables and account for the predictive accuracy of the extant risk assessment instruments warrants further exploration. The basic factor structure needs replication in other samples using confirmatory

factor analysis. If the factor structure can be reliably replicated, future studies should identify the best operationalization of each dimension, determine the average predictive value of each, and calculate the optimum weighting for each. This will allow an integration of different assessment instruments, and has the potential of increasing slightly the overall accuracy. We are also pursuing the application of more sophisticated scaling procedures to these dimensions (Wright & Masters, 1982) and exploring whether non-linear analytic statistics can create improved algorithms for combining these predictors (Bates & Watts, 1988; Breiman et al. 1984; Huet et al., 1996; Lewis, 2000; Ripley, 1996; Seber & Wild, 1989) and developing a more accurate actuarial instrument using the building blocks of existing actuarials. The addition the improved scaling of the best measures of the factors to the subtype predictions described below could also yield benefits.

Second, the role of Need factors should be studied in future samples, and the robustness of their ability to add materially to predictive accuracy investigated. The limited reliability with which Need factors were rated suggests that a preliminary step should involve further specification of the Need rating scales. The present moderate levels of interrater reliability may have been an artifact of limited case files, but it is critical to resolve this both for future research and for practice. The predictive value that the present Need ratings have achieved promises that this endeavor will be rewarded. This appears to be an important route to a meaningful improvement in predictive accuracy.

Third, the pursuit of prediction in specific subgroups also offers promise of improved predictive accuracy. The data indicate a good probability for enhanced prediction among child molesters and possible improvement for rapists by focusing on and including in actuarials, age-preference criteria and specific variables predicting within each subgroup. We are currently

beginning work on this problem. Also of interest, but now only in the beginning stages of development is the further differentiation of child molesters and rapists into generalist and specialist subtypes and the exploration of differentially effective predictors for these subgroups.

Fourth, the complicated effects of age at index offense versus age on discharge should be pursued. The present results, if robust, have major practical implications, but require replication.

Recommendations for Practice

Professionals making decisions would be well advised to continue to apply the structured risk assessment instruments based on static historical factors in making their risk evaluations for whatever purpose they are required. The results of this study confirm that they remained the best validated choices to date for predicting sexual recidivism. We plan to generate tables for each of the risk instruments we have used indicating the false negative and false positive rates at each cutoff for the 3, 10, and 15 year outcome gates. Because the entire sample most closely approximates the sample of individuals evaluated for civil commitment, these rates should provide a guide for the levels of each type of mistake that will be made using each instrument. In applying these actuarial instruments, there are, however, a number of important points that practitioners must keep in mind. First, the prediction level of all the instruments is low. Most of the predictive levels would be described by Sjöstedt & Grann's (2002) interpretative criteria as "marginal," sometimes rising barely into the "moderate" range. Such levels do not warrant high confidence in decisions. Second, typically the instruments are applied to situations in which the base rate of recidivism is low, making it difficult to increase the overall hit rate over simply using the base rates (Meehl & Rosen, 1955). Third, in some situations the cutoffs for decisions are not made to optimize the overall hit rate, but to achieve an alternative aim, such as identifying only the highest risk offenders. Such decision cutoffs are unlikely to yield overall

accuracies that exceed the base rate ,(Knight, 2003; Lloyd & Grove, 2002; Wiggins, 1973). The efficacy of the decisions made using these instruments will hang on the value weights that are placed on each *kind* of error made (Cronbach & Gleser, 1964; Lloyd & Grove, 2002). Fourth, any clinical adjustment that is made by the practitioner to the actuarial is likely to reduce the overall accuracy of the decision (Grove & Meehl, 1996; Grove et al., 2000). Although the actuarials studied in this project provide best guidelines for decision-making, they must be applied by experts with adequate training in applying these instruments and they must be interpreted only by those who appreciate the statistical consequences of their use and the statistical nuances in which these decisions are embedded.

Structured assessments of Need factors should be used cautiously in combination with established empirically derived actuarial instruments. The need for caution arises from two sources. First, that the volume of studies supporting the predictive value of Need ratings is quite small relative to that supporting the value of actuarial instruments. Second, the limited reliability of Need ratings is a serious impediment to effective practical use. Although the promise of Need ratings is great, they will mislead as often as they help if clinicians cannot apply their criteria reliably. Clinicians are therefore advised to draw on the SRA framework or the closely related STABLE rating system, but to work with colleagues to develop consensual Need ratings so that reliability may be enhanced.

Any actuarial that we develop for subsample prediction using the current sample must remain simply a speculative scale until it is replicated on an independent sample. Until that time, it should not be used in decision-making. An extensive literature comparing clinical and actuarial prediction indicates that the clinical adjustment of extant actuarials is substantially more likely to reduce accuracy than to enhance it (Grove & Meehl, 1996; Grove, Zald, Lebow,

Snitz, & Nelson, 2000). Even if the additional subgroup measures that we have identified continue to account for significant additional predictive variance, they should be incorporated into decision-making by the creation of a new actuarial. Clinicians are not likely to be able to incorporate them by making adjustments (Faust, 1986; Garb, 1998).

In light of our data on age, we would not recommend adjusting actuarial risk estimates on the basis of offenders' aging in prison. The present data suggest that inmates do not age out of risk, at least until after the age of 60. Even if some aging adjustment could be empirically validated across samples, such adjustments should be incorporated statistically and not left to clinicians and evaluators. Clearly, much research is needed before this issue will be resolved.

Appendix 1: Actuarial methods to be coded

Detailed, explicit coding manuals exist for all of the instruments that are to be rated in this study. Instead of attaching all of the manuals, which would have entailed hundreds of pages, we have simply summarized the items assessed in each instrument.

The **Violence Risk Assessment Guide (VRAG)** is one of the most widely used actuarial tools for violence offence recidivism (Quinsey et al., 1998). It was developed in Canada, based upon patients detained in secure hospitals between 1965 and 1980, and has been the subject of extensive evaluation that has confirmed its predictive utility. The VRAG contains twelve items, with weighted factors used to assign individuals to one of nine risk categories.

VRAG Items

- 1- Psychopathy Checklist-Revised Score (PCL-R; Hare, 1991)
- 2- Elementary school maladjustment
- 3- DSM diagnosis of any personality disorder
- 4- Age at index offence
- 5- Lived with both parents to the age of 16
- 6- Failure on conditional release
- 7- Criminal history score for non-violent offences
- 8- Marital status
- 9- DSM diagnosis of schizophrenia
- 10-Victim injury
- 11-History of alcohol problems
- 12- Female victim

SORAG Items

- 1- Lived with both parents to the age of 16
- 2- Elementary school maladjustment
- 3- History of alcohol problems
- 4- Marital status
- 5- Criminal history score for non-violent offences
- 6- Failure on conditional release
- 7- Age at index offence
- 8- Victim injury
- 9- Female victim
- 10-DSM diagnosis of any personality disorder
- 11-DSM diagnosis of schizophrenia
- 12-Psychopathy Checklist-Revised Score (PCL-R; Hare, 1991)
- 13- Criminal history score for violent offences
- 14-Number of previous convictions for sexual offenses
- 15-History of sexual offenses only against females under 14
- 16-Phallometric results

The **Rapid Risk Assessment for Sex Offence Recidivism** (RRASOR) is essentially an actuarially based tool that weights a number of key variables in terms of their predictive utility (Hanson, 1997). The initial seven items were based upon a meta-analysis and four were subsequently substantiated as having predictive accuracy for sex offence recidivism. The variables that comprise the tool can be scored to produce an overall risk weighting. The ability of the tool to distinguish between high and low risk has been validated and it has been extensively tested both on both developmental and validation samples.

RRASOR items:

- 1- Prior sex offenses (not including index offenses)
- 2- Age at release (current age)
- 3- Victim gender
- 4- Relationship to victim

STATIC-99 RRASOR and the SACJ-Min have been combined to produce **STATIC 99** (Hanson and Thornton, 1999). Research found that **STATIC 99** outperformed both the RRASOR and SACJ-MIN though the improved prediction achieved was relatively small. **STATIC 99** is a developing tool to which further dynamic risk factors are likely to be added to improve its predictive accuracy.

STATIC-99 items:

- 1- Young
- 2- Ever Lived with an Intimate Partner – 2 Years
- 3- Index Non-Sexual Violence (NSV) – Any Convictions
- 4- Prior Non-Sexual Violence – Any Convictions
- 5- Prior Sex Offences
- 6- Prior Sentencing Dates
- 7- Non-Contact Sex Offences – Any Convictions?
- 8- Any Unrelated Victims?
- 9- Any Stranger Victims?
- 10-Any Male Victims?

Static-2002 Like **Static-99** (Hanson & Thornton, 1999, 2000), **Static-2002** is intended to be a widely applicable risk scale for the prediction of sexual recidivism, which can be coded using commonly available file information.

Static-2002 Items

1. Age at release
2. Sentencing occasions for sexual offences
3. Juvenile arrest for a sexual offence (and convicted as an adult for a separate offence)
4. High rate of sexual offending

5. Any convictions for non-contact sex offences
6. Any male victims
7. Two or more victims < 12 years, one unrelated
8. Any unrelated victims
9. Any stranger victims
10. Arrest/Sentencing Occasions
11. Any breach of conditional release
12. Years free prior to index offense
13. Any convictions for non-sexual violence

The Minnesota Sex Offender Screening Tool (Mn-SOST-R)

Mn-SOST-R Items

Historical/static variables

1. Number of sex/sex-related *convictions* (including current conviction)
2. Length of sexual history
3. Was the offender under any form of supervision when they committed any sex offense for which they were eventually charged or convicted?
4. Was any sex offense committed in public places?
5. Was force or the threat of force ever used to achieve compliance in any sex offense (charged or convicted)?
6. Has any sex offense (charged or convicted) involved 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)
8. Offended against a 13- to 15-year-old victim and the offender was more than five years older than the victim at the time of the offense
9. Was the victim the stranger in any sex-related offense (charged or convicted)
10. Is there evidence of adolescent antisocial behaviour in the file?
11. Substantial history of drug or alcohol abuse (12 months prior to instant offense or revocation)
12. Employment history

Dynamic/institutional variables

13. Discipline history while incarcerated (does not include discipline for failure to follow treatment directives)
14. Chemical dependency treatment while incarcerated or on release
15. Sex offender treatment history while incarcerated or on release
16. Age of the offender at time of release

The Adult Sex Offender Assessment Protocol (A-SOAP-II)

The Adult Sex Offender Assessment Protocol, A-SOAP, was developed in 1994 when I was at the Joseph J. Peters Institute in Philadelphia. It is an archival risk assessment scale for adult sexual offenders. There are four rationally derived scales, two static and two dynamic.

A-SOAP-II items

Sexual Drive/Preoccupation Scale

1. Prior Legally Charged Sex Offenses
2. Number of Sexual Abuse Victims
3. Male Child Victim
4. Duration of Sex Offense History
5. Degree of Planning in Sexual Offense/s
6. Sexualized Aggression
7. Sexual Drive and Preoccupation
8. Sexual Victimization History

Impulsive, Antisocial Behavior Scale

9. Juvenile Antisocial Behavior
10. Ever Charged/Arrested Before Age 16
11. Adult Antisocial Behavior
12. Pervasive Anger
13. Multiple Types of Offenses
14. Impulsive Lifestyle
15. Physical Assault History and/or Exposure to Family Violence

Intervention Scale

16. Accepting Responsibility for Offense/s
17. Internal Motivation for Change
18. Understands Risk Factors/Applies Strategies
19. Empathy²
20. Remorse and Guilt
21. Cognitive Distortions
22. Quality of Relationships

Community Stability/Adjustment Scale

23. Intimacy Needs
24. Management of Anger
25. Work Stability
26. Support Systems

Risk MATRIX 2000 (Hanson and Thornton, 2000) represents an important improvement on the Structured Anchored Clinical Judgement (SACJ) as it provides for greater accuracy and refinement in the identification of high risk offenders, and offers two versions, one for sex offenders and one for violent offenders. Although the tool has not yet been subject to extensive published evaluations, it has been validated retrospectively. Development and validation of both versions have, however, been undertaken with male offenders (and often male prisoners) and they may have a limited transferability to other groups.

Risk MATRIX/S Scale items

Step One Factors

- Number of Occasions Sentenced for a sex offense, differentiated into 4 levels.
- Number of Occasions Sentenced for Any Criminal Offense, differentiated into 2 levels
- Age on release, differentiated into three levels

Step Two Factors

- Male victim of a sexual offense.
- Stranger victim of a sexual offense.
- Single (Never Married)
- Non-Contact Sex Offense

Risk MATRIX/V Scale items

- Age.
- Violent Appearances.
- Burglary.

Risk MATRIX/C Scale items

Combines the S and V ratings.

Sexual Violence Risk-20 (SVR-20) - is a 20-item instrument that covers a range of static and dynamic factors and includes items aimed at informing risk management plans (Boer et al., 1997). Some of the items in the tool are drawn directly from the HCR-20.

SVR-20 items:

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 offences
- 10-Past non-violent offences
- 11-Past supervision failure

Sexual offence:

- 12-High density sex offences
- 13-Multiple sex offence types
- 14-Physical harm to victim(s)
- 15-Use of weapons or threats of death in sex offenses
- 16-Escalation in frequency or severity of sex offenses
- 17-Extreme minimization or denial of sex offenses
- 18-Attitudes that support or condone sex offenses

Future plans:

- 19-Lacks realistic plans
- 20-Negative attitude towards intervention

CODING SRA FACTORS

After an initial inspection of the case files it was decided that no attempt would be made to rate factors from the Distorted Attitudes domain as the clinicians who made entries in the case files did not seem sensitive to modern concepts of cognitive distortions. Similarly, the Callousness factor was not included as it was thought that the case files did not have adequate information to code this validly. Additionally some SRA factors were already present in the other scales that were coded so there was no need to provide separate coding instructions for them. The main example of this is factor two from the PCL-R, which is seen as part of the Self-Management domain in SRA.

Once the data had been collected, the SRA factors defined below, and PCL-R factor two, were correlated with conceptually related, potentially dynamic, factors from the other scales. Where a factor correlated substantially with conceptually related SRA marker variables it was combined with the other SRA variables in creating the SRA-Need score. This selection and combination of items was made prior to examining the relation of the scores to recidivism.

The additional variables were:

Sexual Drive and Preoccupation from the ASOAP
 Pervasive Anger from the ASOAP
 Impulsive Lifestyle from the ASOAP

Sexual Deviation from the SVR-20

These variables were already coded into three levels (0,1,2) similar to the way the SRA marker variables had been coded. PCL-R factor two was reduced to three levels by using cutting lines that coded it as 0 for raw scores below the sample mean, coded it as 2 for raw scores more than one standard deviation above the sample mean, and coded 1 if the raw scores were between these two levels.

The SRA-Need score was the sum of these eleven factors.

Child-Preference

Concept

Child-preference means a stronger sexual response to children than adults.

In scoring this item Child here means prepubescent females (12 or under if no other information), or, prepubescent or young teenage males. Adult refers to someone aged at least 18.

Coding

Generally Applies = 2

Sufficient evidence for a 2 includes any of the following.

- 3 or more child victims with offending against them spanning more than 6 months.
- All of the four following criteria need to be met for this bullet to apply:
 - male victim under 14,
 - unrelated victim under 14,
 - at least 2 victims under 14,
 - at least one victim under 11.
- Self-report of sexual fantasies about or attraction to children sustained¹ for more than six months.
- 2 or more child victims and an absence of evidence for a sexual interest in adults.
- Found in possession of child-pornography; collects pictures of children etc.

Partially Applies = 1

Any child victims or self-report of sexual attraction to children but fails to meet criteria for a 2.

¹ Being sustained for more than 6 months does not mean there must be evidence of the fantasies / interests being continuously present for more than six months. This principle applies generally in this guide.

Does Not Apply = 0

No child victims and no self-report of sexual attraction to children combined with evidence for a sexual interest in adults.

Note: If the offender was under the age of 18 when offenses were committed, do not count offenses as “against a child” or against someone “under 14” unless there was at least 5 years difference in age between the perpetrator and the victim.

Sexualized Violence**Concept**

Preferring coerced sex over consenting sex

OR

Significant sexual response to someone else’s pain, terror, or humiliation.

The common element is abusive control/domination having become a significant source of sexual arousal. Note that the mere fact that a sexual assault is painful/terrifying or humiliating for the victim does not mean that this factor applies to the perpetrator.

Coding**Generally Applies = 2**

Sufficient evidence for a 2 includes any of the following.

- At least 3 victims with offenses spread over more than 6 months where there was behavior that was liable to cause humiliation, pain or terror beyond that needed to secure victim compliance.
- Self-report of sustained rape fantasies (sustained means evidence that these fantasies spanning more than 6 months).

Partially Applies = 1

Criteria for a 2 not met but at least one offense in which there was behavior that was liable to cause humiliation, pain or terror beyond that needed to secure victim compliance or report of rape fantasies but no evidence that they were sustained.

Does Not Apply = 0

No offenses in which there was behavior that was liable to cause humiliation, pain or terror beyond that needed to secure victim compliance and no self-report of rape fantasies.

Sexual Pre-occupation

Concept

An intense interest in sex; much behavior is sexually motivated.

Coding***Generally Applies = 2***

Sufficient evidence for coding a 2 includes at least three of the following being true.

- Repeated casual sex; repeated affairs while in marital type relationship; large number of different sexual partners.
- Repeated use of pornography.
- Repeated sexual behavior that is out of place such as making sexual comments in non-sexual situations.
- Three or more distinct sex acts within any offense event², repeatedly masturbating more than once a day, high rate of sex acts with others (regularly more than 5 times a week), repeated experience of deviant sexual fantasies.
- Meets DSM criteria for at least 2 different paraphilias³.

Note: any exemplar after a bullet means the bullet applies.

Partially Applies = 1

Does not meet the criteria for a 2 but at least one of the above bullets applies.

Does Not Apply = 0

None of the above bullets applies.

Lack of Emotionally Intimate Adult Relationships**Concept**

This refers to the absence of emotionally intimate marital type relationships.

² Count the following as distinct sex acts: oral sex, anal sex, vaginal sex, penetration with object/finger. If the same sex act is repeated within the offense event with a gap in time between repetitions, count each repetition as a separate act.

³ Paraphilias are recurrent sexual fantasies, urges or behaviors involving non-human objects, suffering or humiliation of oneself or one's partner, or children or other non-consenting persons that occur over a period of at least 6 months. Commonly identified paraphilias are Exhibitionism, Fetishism, Frotteurism, pedophilia, sexual masochism, sexual sadism, and tranvestic fetishism. Additionally, Paraphilia NOS (Non-Consent) is often identified in offenders. See the DSM IV manual.

Marital type relationships involve two adults living together as lovers, sharing a household, sharing bills, and making a life together for at least two years.

A marital-type relationship can be seen as lacking emotional intimacy if frequent fights, domestic violence, repeated infidelity, or emotional distance between the partners mar it.

Coding

Generally Applies = 2

Any of the following would be sufficient evidence for a 2.

- Never had a marital type relationship that was sustained for at least two years.
- Had such a relationship for at least two years but it was marred by frequent fights, repeated infidelity, or by domestic violence.
- Had such a relationship for at least two years but relationships of this kind were marred by emotional distance in which partners were unwilling to talk to each other about what really mattered to them.

Note: Committing sexual offenses while in marital type relationship would count as infidelity.

Partially Applies = 1

As an adult, had a marital type relationship that lasted at least two years but some evidence of relationship problems that does not meet the criteria for a 2. Also score a 1 if no adult marital type relationships because he has been consistently in institutions since the age of 18.

Does Not Apply

As an adult, had a marital type relationship that lasted at least two years and it was not marred by any of the problems indicated above.

Emotional Congruence with Children

Concept

This refers to finding it easier to relate to children than to adults, to preferring the company and companionship of children to that of adults.

Coding

Generally Applies = 2

To score a 2 both of the following should apply.

- Repeatedly making/seeking social relationship or friendships with children; including spending substantial time in a “kindly” grooming relationship with children prior to offending.
- Difficulty, awkwardness or avoidance of friendships or romantic relationships with adults.

Here “grooming” does not include threats or intimidation,

Partially Applies = 1

The first of the two criteria above apply but the second does not (that is, he seeks social relationships with children but can relate to adult as well).

Does Not Apply = 0

Neither of the two criteria above apply.

Grievance Thinking

Concept

Easily feels wronged; suspicious; ruminates angrily; tends not to see or accept other’s point of view

Coding

Generally Applies = 2

Positive evidence for at least two of the following features over a period that spans more than six months.

- Easily feels wronged; suspicious of others
- Ruminates angrily
- Tends not to see or accept other’s point of view

Note that a bullet applies if any of the features listed by it applies

Partially Applies = 1

Does not meet the criteria for a 2 but positive evidence for at least one of these features.

Does not apply = 0

No positive evidence for any of these features or weak evidence for them that is contradicted by strong evidence of contrary functioning.

Dysfunctional Coping

Concept

Reacts to stress or problems in an impulsive/reckless way; an absence of effective systematic problem-solving, or problem-solving abilities disrupted by emotion-focused or avoidant coping⁴.

Coding

Generally Applies = 2

At least two positive pieces of evidence for reacting to stress or problems in an impulsive/reckless way that span more than six months combined with an absence of evidence for good self-control and good problem-solving abilities.

Note: problem-solving abilities that are regularly disrupted by emotional reactions do not count as “good problem-solving abilities”.

Partially Applies = 1

Fails to meet the criteria for a 2 but some evidence for reacting to stress or problems in an impulsive/reckless way. OR no such evidence but no positive evidence of good self-control and problem-solving abilities

Does Not Apply = 0

No evidence for reacting to stress or problems in an impulsive/reckless way and positive evidence of good self-control and problem-solving abilities.

⁴ Emotion-focused coping is responding to stress or problems by emotionally dwelling on the negative aspects of the situation rather than thinking about the matter in a practical and proportionate way. Avoidant-coping is responding to stress or problems by trying to distract yourself from the issue rather than thinking about it in a practical and proportionate way.

Appendix 2: Additional measures to be coded

DSM-IV Diagnoses

- Axis 1: Clinical disorders
- Axis 2: Personality disorders

PCL-R Items

The Psychopathy Check List-Revised (PCL-R) and its derivatives (the PCL:YV for adolescents and the PCL:SV 'screening version') constitute clinical construct rating scales used in semi-structured interview (Hare, 1991). It involves rating 20 items on a 3 point scale divided into three broad categories: interpersonal/affective, social deviance and 'additional items'. It has also been shown to be a highly reliable tool when used by well-trained assessors.

Factor 1

1. Glibness/superficial charm
2. Grandiose sense of self-worth
3. Pathological lying
4. Conning/manipulative
5. Lack of remorse or guilt
6. Shallow affect
7. Callous/lack of empathy
8. Failure to accept responsibility for own actions

Factor 2

9. Need for stimulation/-proneness to boredom
10. Parasitic lifestyle
11. Poor behavioral controls
12. Early behavioral problems
13. Lack of realistic, long-term goals
14. Impulsivity
15. Irresponsibility
16. Juvenile delinquency
17. Revocation of conditional release

Additional items

18. Promiscuous sexual behavior
19. Many short-term marital relationships
20. Criminal versatility

Conduct reports for:

- Major
- Minor
- Violent
- Non-violent
- Sexual of any kind
- Non-Sexual
- Sexual Assaults

Treatment Participation:

- Drop-out < 12 months
- Drop-out = or > 12 months
- Participating but treatment incomplete at release
- Treatment completed prior to release

Long-term psychological risk factors from the SRA framework**Sexual interests domain**

- Pedophilic preference
- Coercion preference or sadism
- Sexual pre-occupation

Distorted Attitudes domain

- Child-abuse supportive beliefs
- Rape-minimization
- Rape as Justified (e.g. Revenge/Punishment/Re-affirmation of control)
- Deceitful Women
- Sexual Entitlement

Socio-affective Functioning domain

- Inadequacy (low self-esteem; lonely; external locus of control; submissive)
- Lack of Emotionally Intimate Relationships with adults
- Callous / Shallow Emotions (from PCL-R)
- Grievance Thinking (Ruminates anger; suspicious of others; poor perspective taking; typically feels has been wronged and wants revenge)

Self-Management domain

- Life-style Impulsiveness (from PCL-R , basically factor 2)
- Poor Cognitive-Problem-Solving
- Poor Control of Emotional impulses

Appendix 3: Select MTC Coding Dictionary Scales and MTC:R3 and MTC:CM3 Scales

MTC Coding Dictionary Items:

Adult Sex Crimes—For rapists the total number of sexual crimes charged in which the victim ≥ 16 , and for child molesters the total number of sexual crimes charged in which the victim ≤ 15 .

Male Victim Choice - the listing of a male victim (≤ 15 years of age) in any charge or conviction or as admitted by the offender in treatment.

MTC Coding Dictionary Scales and the Items They Comprise:

Scale	Items	α
<i>Impulsivity</i>	Impulsive (0, 1, 2)	.66
	Rebellious (0, 1, 2)	
	Unstable employment (0, 1)	
	financial irresponsibility (0, 1)	
	reckless behavior (0, 1)	
<i>Paraphilias</i>	Exhibitionism (0, 1, 2)	.66
	Masturbation (0, 1, 2)	
	Fetishism (0, 1, 2)	
	Transvestism (0, 1, 2)	
	Voyeurism (0, 1, 2)	
<i>Social Isolation</i>	Passive (0, 1, 2)	.80
	Isolated (0, 1, 2)	
	Seclusive (0, 1, 2)	
	Shy (0, 1, 2)	
	peer relationship problems (0, 1, 2)	

MTC:CM3 and MTC:R3 Scales:

Level of Fixation on Children (CM3) – (interrater reliability for consensed judgment = .83)

The “level of fixation” decision attempts to assess the strength of an offender’s pedophilic interest (i.e., the extent to which children are a major focus of the offender’s thought and attention). If unequivocal, direct evidence (e.g., direct report by the offender of the presence, nature, and duration of fantasies about children) is available, indicating that children have been a central focus of the offender’s sexual and interpersonal fantasies and cognitions for a protracted period (at least six months), rate the offender as having high fixation.

In the absence of direct evidence about the offender's fantasies and cognitions, the following criteria should be used to guide this decision. Because the clinical files that were our data source often lacked the more direct evidence described above, these supplementary criteria were the bases for most of the fixation judgments in the present study.

Low fixation. The offender is considered to be "low fixated," if he is over 20 years old, and all of his sexual encounters with children (both charged and uncharged) occurred within a six-month period (here "children" is interpreted to mean youngsters at least 5 years younger than the offender. If an offender is 20 years old and younger, and all of his offenses occurred within a six-month period, he may be "high fixated," if he fits Criteria B2 or B3 under "high fixation" below or there is evidence of serious deficiencies in his peer relationships in adolescence (as evidenced by lack of age-appropriate acquaintances or of considerable time spent with preadolescent children).

High fixation. An offender is considered high fixated if:

- (A) He does not fit the criteria for "low fixation" and/or
- (B) Any of the following are present:
 - (1) There is evidence of three or more sexual encounters with children, and the time period between the first and third encounter was greater than six months. These encounters may be with a single victim over many incidents, and should not be limited to charged offenses. They do not include incest offenses.
 - (2) There is evidence that the offender has had enduring relationships with children (excluding parental contact). This includes sexual and nonsexual and professional and nonprofessional contact.
 - (3) The offender has initiated contact with children in numerous situations over his lifetime.

Pervasive Anger (R3) (Interrater reliability for consensed rating = .73)

This scale is the sum of the following five items, each of which is given a 0 (absent) or 1 (present) rating—

1. The offender or others characterized him as an angry person who easily loses his temper and is likely to get in trouble because of his hostility. This anger is directed at multiple targets and appears in multiple situations. It does not appear to be exclusively focused at particular people or specific issues, or to occur in isolated situations.
2. The offender has shown a consistent pattern of verbal aggression against both males and females, manifesting angry verbal attacks against peers and authority figures on multiple occasions. Do not, however, rate as present if the offender is only angry at authority figures.
3. Either the offender has assaulted males, and these assaults against males appear to have been motivated by anger or hostility, rather than by any sexual intent, or the offender has frequently (on more than two occasions) gotten into physical fights with males.

4. The offender reports *preoccupation* with aggressive fantasies that include thoughts of beating, killing, torturing, or mutilating others. These fantasies clearly involve inflicting pain or putting someone in excessive fear or discomfort.
5. The offender reports or is reported to have been cruel to animals, which includes having beaten, tortured, mutilated, or killed them. His treatment of animals must clearly have involved inflicting pain or killing them.

Offense Planning (R3) (*Interrater reliability for averaged ratings = .54*)

The maximum offense planning score was the highest level of planning that any one offense achieved for an offender on a 0 to 4 scale for the anchors that are described below—

Detailed Planning (4) -- The offense was planned in detail and a particular victim or type of victim was sought. This includes, but is not limited to, scripted offenses, in which the *modus operandi* of the offense follows an apparent "script" that seems to be related to cognitions and fantasies that precede the offense.

High Moderate Planning (3) -- In this type of offense the high consistency of the offender's behaviors across offenses or particular behaviors like observing a particular victim on several occasions before the assault indicate that considerable forethought and planning preceded the offense.

Moderate Planning (2) -- In this type of offense, before the victim was encountered, the offender had conceived of the idea of committing a sexual offense. That is, the offender does not simply set out on impulse, or with a vague intention to seek sexual gratification. Although he may not have a particular victim in mind, it is clear from the kinds of equipment he takes with him, the place to which he chooses to go, and his behaviors before and during the offense that coercive sexual behavior was intended before a victim was encountered.

Low Moderate Planning (1) -- In this type of offense the encounter with the victim plays only a moderate role. There is evidence from some aspect of the assault, whether a vague similarity in *modus operandi* to previous assaults, or a similarity in the locations of assaults or the approach to the victim, that suggests that there was at least a vague intention to force a victim into sexual compliance prior to encountering the victim. The crime of an offender who puts himself in circumstances in which he may encounter a victim (e.g., cruising in his car in particular locations) can be characterized as "low moderate planning."

Impulsive Offense (0) -- In this type of offense the encounter with the victim appears to have played an important role in eliciting the offense. For example, the offense occurred during another crime, in which a victim was unexpectedly encountered, and was raped because of convenient availability. In cases in which the offender knows the victim, the offense can be considered impulsive, even if the offender had the intention of sexually engaging, but not raping, the victim before the assault. In such cases the rape should

appear to have occurred when the offender's sexual advances were thwarted, and the rape resulted from his failure to inhibit his sexual/ aggressive impulses.

Juvenile Unsocialized Aggression (R3) (Interrater reliability for consensed rating = .91)

This scale is the sum of the following six items, each of which is given a 0 or 1 rating in accordance with the criteria described—

1. Problems in grammar school (grades K - 6)
 - 0 = No problems or only minor attendance/discipline problems
 - 1 = Moderate to severe behavior problems (disciplinary and/or 1 attendance problems, including chronic truancy)
2. Problems in junior high school (grades 7 - 9) -- Coded the same as item #1
3. Total number of non-sexual victimless offenses prior to 16th birthday
 - 0 = none or only one
 - 1 = two or more
4. Running away prior to 17th birthday
 - 0 = no
 - 1 = yes
5. Vandalism and destruction of property prior to 16th birthday
 - 0 = no evidence
 - 1 = yes, evidence for intentional destruction of property
6. Involved in fights prior to 16th birthday
 - 0 = no evidence
 - 1 = yes, evidence for involvement in fights on more than one occasion (exclude fights with siblings)

Adult Unsocialized (R3) Aggression (Interrater reliability for consensed rating = .90)

This scale is the sum of the following eight items, each of which is given a 0 or 1 rating in accordance with the criteria described—

1. History of non-prescription drug use
 - 0 = no evidence
 - 1 = yes, evidence for use of illegal or "street" drugs
2. Vandalism and/or destruction of property at age 16 or older
 - 0 = no evidence
 - 1 = yes, evidence for intentional destruction of property
3. Fighting at age 16 or older
 - 0 = no evidence
 - 1 = yes, evidence for involvement in fights on more than one occasion
4. Assaultive Offenses
 - 0 = no evidence
 - 1 = one or more arrests (other than sexual crimes) for any offenses in which he was physically assaultive
5. Unsocialized Aggression

- 0 = no more than frequent mild aggression (e.g., spats/arguments, verbal aggression)
- 1 = at least occasional moderate aggression that is manifest physically (e.g., fights, brawls, or minor assaults on two or more occasions, excluding sexual crimes)
6. Conduct/behavioral charges
- 0 = none
- 1 = one or more charges for drunk, disorderly, disturbing the peace, defacing property, etc.
7. Owned a manufactured weapon -- do not count a knife unless the subject used it specifically as a weapon, but do count brass knuckles
- 0 = no
- 1 = yes
8. Relation between alcohol use and antisocial behavior
- 0 = it is atypical that acting out occurs during or after drinking, or such behavior is infrequent
- 1 = acting out usually occurs during or after drinking

Sadism (R3) (*Interrater reliability for consensed rating = .80*)

This scale is the sum of the following eight items, each of which is given a 0 or 1 rating—

- (1) The offender reports *preoccupation* that is both sexual and aggressive in nature (for instance, sexual fantasies that include thoughts of beating, raping, torturing, or killing). These fantasies may involve more detailed scenes or scripts in which inflicting pain or putting the victim in excessive fear are *key and clear* intents. Fantasies of raping without evidence of such direct intentions of causing the victim pain and/or fear are not sufficient for this criterion.
- (2) The victim's pain, fear, or discomfort appears to facilitate sexual arousal and/or lead to ejaculation. Consistent with the general description of sadism, there should be no evidence that the offender lost his erection or failed to ejaculate while he was assaulting the victim, unless the assault was interrupted by some external event.
- (3) There is clear evidence of sham sadism in the sexual offenses, which need not be violent and may be limited to such behavior as sham whipping or bondage.
- (4) There is clear evidence of sham sadism in the offender's consensual sexual relationships, which need not be violent and may be limited to such behavior as sham whipping or bondage.
- (5) In his consensual sexual relationships there is clear evidence of overt sadism, indicated by the presence in these relationships of Item #6 (below) from Category A or two or more of the behaviors from Category B of the Sadism Criteria.
- (6) The violence in the offense(s) is ritualized, indicating an underlying fantasy or script (e.g., there is repetition of a particular sequence of acts or there is an ordered sequence that was clearly not conceived on the spot).

- (7) The offender has intercourse with the victim after the victim has been killed.
- (8) The offender mutilates the victim's erogenous zones (e.g., vagina, penis [for male victims], breasts, anus, buttocks, etc.) after the victim is dead.

References

- Andrews, D. A. & Bonta, J. L. (1995). *The Level of Service Inventory: Revised Users Manual*. NY, North Towanda: Multi-Health Systems.
- Barbaree, H. E. (1997). Evaluating treatment efficacy with sexual offenders: The insensitivity of recidivism studies to treatment effects. *Sexual Abuse: A Journal of Research and Treatment*, 9, 111-128..
- Barbaree, H. E. (2003). Risk assessment: Discussion of the session. In R. A. Prentky, E. Janus, & M. Seto (Eds.), *Sexual coercion: Understanding and management*, pp. 238-240. New York: New York Academy of Sciences.
- Barbaree, H. E., Blanchard, R., & Langton, C. M. (2003). The development of sexual aggression through the life span: The effect of age on sexual arousal and recidivism among sex offenders. In R. Prentky, E. Janus, & M. Seto (Eds.), *Understanding and managing sexually coercive behavior* (pp. 59-71). Annals of the New York Academy of Sciences, Vol. 989. New York: New York Academy of Sciences.
- Barbaree, H. E., Seto, M. C., Langton, C., & Peacock, E. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice & Behavior*, 28(4), 490-521.
- Barbaree, H. E., Langton, C. M., & Peacock, E. J. (in press) The factor structure of static actuarial items: Its relation to prediction. *Sexual Abuse: A Journal of Research and Treatment*.
- Bard, L. A., Carter, D. L., Cerce, D. D., Knight, R. A., Rosenberg, R., Schneider, B. (1987). A descriptive study of rapists and child molesters: Developmental, clinical, and criminal characteristics. *Behavioral Sciences and the Law*, 5, 203-220.

- Bates, D.M., & Watts, D.G. (1988). *Nonlinear regression analysis and its applications*. New York: John Wiley & Sons.
- Beech, A. R., Fisher, D. D., & Thornton, D. (2003). Risk assessment of sex offenders. *Professional Psychology: Research and Practice, 34*, 339-352.
- Beech, A. R., & Ward, T. (2006). Risk assessment in the 21st century: Towards an integrative model of risk. *The Forum*, Summer.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk – 20: Professional Guidelines for Assessing Risk of Sexual Violence*. Vancouver, B.C.: The Institute Against Family Violence.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.
- Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment, 13*, 171-188.
- Craig, L. A., Beech, A. R., & Browne, K. D. (2006). Evaluating the predictive accuracy of sex offender risk assessment measures on UK samples: A cross-validation of the Risk Matrix 2000 Scales. *Sex Offender Treatment, 1*, 1-16.
- Craig, L. A., Thornton, D., Beech, A., & Browne, K. D. (in press). The relationship of statistical and psychological risk markers to sexual reconviction in child molesters. *Criminal Justice and Behavior*.
- Cronbach, L. J., & Gleser, G. C. (1964). *Psychological tests and personnel decision*. (2nd Ed.) Urbana: University of Illinois Press.
- Daversa, M. T., & Knight, R. A. (in press). A structural examination of the predictors of sexual coercion against children in adolescent sexual offenders. *Criminal Justice and Behavior*.

- Daversa, M. T., Sitnikov, L., & Knight, R. A. (2005, October). *Investigating the relation of secondary psychopathy and borderline personality features in a sample of sexual offenders*. Poster presented at the 20th Annual Meeting of the Society for Research in Psychopathology, Miami, Florida.
- DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L., (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach, *Biometrics*, 44, 837-845.
- Doren, D. M., & Epperson, D. L. (2001). Great analysis, but problematic assumptions: A critique of Janus and Meehl (1997). *Sexual Abuse: A Journal of Research and Treatment*, 13, 45-51.
- Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology*, 67, 917-930.
- Douglas, K. S., Vincent, G. M., & Edens, J. F. (2006). Risk for criminal recidivism: The role of psychopathy. In C. Patrick (Ed.), *Handbook of psychopathy*. pp. 533-554. New York: Wiley.
- Douglas, K. S., & Webster, C. D. (1999). The HCR-20 violence risk assessment scheme: Concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior*, 26, 3-19.
- Epperson, D. L. (2000). *Description of the cross-validation sample*. Retrieved from <http://psych-server.iastate.edu/faculty/epperson/crossvalidationupdate.html>.

Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (1998).

Minnesota Sex Offender Screening Tool – Revised (MnSOST-R): Development, performance, and recommended risk level cut scores. Available at:

http://www.psychology.iastate.edu/faculty/epperson/mnsost_download.htm.

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice*, 17, 420-430.

Garb, H. N. (1998). *Studying the clinician*. Washington, DC: American Psychological Association.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.

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

Hair, J. F., & Black, W. C. (2002). Cluster analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 147-205). Washington, DC: American Psychological Association.

Hanley, J.A. and McNeil, B.J., (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases, *Radiology*, 148, 839-843.

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. Department of the Solicitor General of Canada, Ottawa, Ontario. (Available at www.sgc.gc.ca)

Hanson, R. K. (2006, June). A Framework for Violence Risk Assessment: Static, Stable and Acute Factors. Canadian Psychological Association, Calgary, Alberta

- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, **66** (2), 348-362.
- Hanson, R. K., & Morton-Bourgon, K. E. (2004). *Predictors of sexual recidivism: An updated meta-analysis*. Available at: http://www.psepc.gc.ca/publications/corrections/pdf/200402_e.pdf.
- Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders* (User report 1999-02). Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, *24*(1), 119-136.
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002. User Report 2003-01*. Ottawa: Department of the Solicitor General of Canada.
- Hare, R. D. (1991). *Manual for the revised Psychopathy Checklist*. Toronto: Multi-Health Systems.
- Hare, R. D., & Neumann, C. S. (2006). The PCL-R assessment of psychopathy: Development, structural properties, and new directions. In C. Patrick (Ed.), *Handbook of psychopathy*. pp. 58-88. New York: Wiley.
- Harris, D. A., Knight, R. A., Smallbone, S., & Dennison, S. (2006, July) *Offence specialisation and versatility in men convicted of sexual offences*. Paper presented at the 20th Annual Meeting of the International Family Violence and Child Victimization Conference, Portsmouth, New Hampshire.
- Harris, C. M., Kaylan, A. R., & Maltz, M. D. (1981). Recent advances in the statistics of recidivism measurement. In J. A. Fox (Ed.), *Models in quantitative criminology* (pp. 61-79). New York: Academic Press.

- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315-335.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). *A multi-site comparison of actuarial risk instruments for sex offenders*. Submitted for publication.
- Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology, 3*, 141-172.
- Huet, S., Bouvier, A., Gruet, M.-A. & Jolie, E. (1996). *Statistical tools for nonlinear regression: A practical guide with S-PLUS examples*. New York: Springer-Verlag.
- Janus, E. S., & Meehl, P. E. (1997). Assessing the legal standard for prediction of dangerousness in sex offender commitment proceedings. *Psychology, Public Policy, and Law, 3*, 33-64.
- Jenkins, W. O., Barton, M. C., de Valera, E. K., DeVine, M. D., Witherspoon, A. D., & Muller, J. B. (1972). *The measurement and prediction of criminal behavior and recidivism: The environmental deprivation scale (EDS) and the maladaptive behavior record (MBR)*. Rehabilitation Research Foundation, Contract No. 82-01-69-06, Manpower Administrations, U.S. Dept. of Labor.
- Jenkins, W. O., DeVine, M. D., de Valera, E. K., Muller, J. B., Nichols, W. C., & Ray, S. O. (1975). *The behavioral demography of the young adult male offender*. Rehabilitation Research Foundation, Manpower Administration, U.S. Dept. of Labor.
- Knight, R. A. (1999). Validation of a typology for rapists. *Journal of Interpersonal Violence, 14*, 297-323.

- Knight, R. A. (2003). Risk assessment: Discussion of the session. In R. A. Prentky, E. Janus, & M. Seto (Eds.), *Sexual coercion: Understanding and management*, pp. 241-245. New York: New York Academy of Sciences.
- Knight, R. A., Carter, D. L., & Prentky, R. A. (1989). A system for the classification of child molesters: Reliability and application. *Journal of Interpersonal Violence* 4, 3-23.
- Knight, R. A., & Guay, J. P. (2006). The role of psychopathy in sexual offenders against women. In C. Patrick (Ed.), *Handbook of psychopathy*. pp. 512-532. New York: Wiley.
- Knight, R. A., & Prentky, R. A. (1990). Classifying sexual offenders: The development and corroboration of taxonomic models. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *The handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 23-52). New York: Plenum Press.
- Knight, R. A., Rosenberg, R., & Schneider, B. (1985). Classification of sexual offenders: Perspectives, methods, and validation. In A. W. Burgess (Ed.), *Research handbook on rape and sexual assault* (pp. 222-293). New York: Garland Publishing Inc.
- Knight, R. A., & Sims-Knight, J. E. (2003). Developmental antecedents of sexual coercion against women: Testing of alternative hypotheses with structural equation modeling In R. A. Prentky, E. Janus, & M. Seto (Eds.), *Sexual coercion: Understanding and management*, pp. 72-85. New York: New York Academy of Sciences.
- Knight, R. A., & Sims-Knight, J. E. (2004). Testing an etiological model for male juvenile sexual offending against females. *Journal of Child Sexual Abuse*, 13, 33-55.
- Langton, C. M. (2003). Contrasting approaches to risk assessment with adult male sexual offenders: An evaluation of recidivism prediction schemes and the utility of

- supplementary clinical information for enhancing predictive accuracy. Unpublished doctoral thesis, University of Toronto, Toronto, Canada.
- Lewis, R. J. (2000). *An introduction to classification and regression tree (CART) analysis*. Paper presented at the Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California.
- Lloyd, M. D., & Grove, W. M. (2002). *The uselessness of the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) in commitment decisions*. Manuscript submitted for publication.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychological Bulletin*, *52*, 194-216.
- Metz, C. E. (1998). ROCKIT (Version 0.9.1b) [Computer software]. Chicago: University of Chicago. Retrieved from <http://www-radiology.uchicago.edu/krl/toppage11.htm>.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*, 783-792.
- Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines*. Ottawa: Solicitor-General of Canada.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill Book Company.
- Poythress, N. G., & Skeem, J. L. (2006). Disaggregating psychopathy. In C. Patrick (Ed.), *Handbook of psychopathy*, pp. 172-192. New York: Wiley.
- Prentky, R. A., Harris, B., Frizzel, K., & Righthand, S. (2000). An actuarial procedure for assessing risk with juvenile sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, *12*, 71-93.

- Prentky, R. A., & Knight, R. A. (1991). Identifying critical dimensions for discriminating among rapists. *Journal of Consulting and Clinical Psychology, 59*, 643-661.
- Prentky, R. A., Knight, R. A., & Lee, A. F. S. (1997). Risk factors associated with recidivism among extrafamilial child molesters. *Journal of Consulting and Clinical Psychology, 65*, 141-149.
- Prentky, R. A., & Lee, A. F. S. (in press). Effect of age-at-release on long term sexual re-offense rates in civilly committed sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*.
- Prentky, R. A., & Righthand, S. (2003). *Adult Sex Offender Assessment Protocol: Manual*. Bridgewater, MA: Justice Resource Institute.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C.A. (1998). *Violent offenders: Appraising and managing risk*. Washington, D.C.: American Psychological Association.
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.
- Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.
- Roberts, C. F., Doren, D. M., & Thornton, D. (2002). Dimensions associated with assessments of sex offender recidivism risk. *Criminal Justice and Behavior, 29*, 569-589.
- Schmidt, P., & White, A. D. (1988). Predicting recidivism using survival models. New York: Springer-Verlag.
- Seber, G.A.F. & Wild, C.J. (1989). *Nonlinear regression*. New York: John Wiley & Sons.
- Serin, R. C., Mailloux, D. L., & Malcolm, P. B. (2001). Psychopathy, deviant sexual arousal, and recidivism among sexual offenders. *Journal of Interpersonal Violence, 16*, 234-246.

- Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment, 17*, 156-167.
- Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health, 1*, 179-183.
- Swets, J. A. (1986). Forms of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement performance. *Psychological Bulletin, 99*, 181-198.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest: A Journal of the American Psychological Society, 1*, 1-26.
- Thornton, D. (2002). *Scoring Guide For Risk Matrix 2000.3*. Madison, Wisconsin:
- Thornton, D. & Beech, A. R. (2002, October). *Integrating statistical and psychological factors through the structured risk assessment model*. Poster presented at the 21st Association for the treatment of Sexual Abusers Conference, Montreal, Canada.
- Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.) *Understanding and managing sexually coercive behavior*. (pp. 225-235). New York: Annals of the New York Academy of Sciences, Vol. 989.
- Vida, S. (2006) *Accumetric: Test performance analysis manual*, Montréal, Québec, Accumetric corporation.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley Publishing Company.

Wright, R. E. (2000). Survival analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 363-407). Washington, D.C.: American Psychological Association.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.

Author Note

This research was supported by research grant 2003WGBX1002 from the National Institute of Justice.

Correspondence concerning this article should be addressed to Raymond Knight, Ph.D., Department of Psychology, MS 062, Brandeis University, Waltham, Massachusetts 02454-9110. Electronic mail may be sent to knight2@brandeis.edu.

We would like to recognize the contribution to this project of the following project staff, without whose dedication and skill nothing would have been possible.

- LuAnn Daniels coordinated the risk assessment coding team and also coded a substantial number cases herself.
- Danielle Harris applied the Massachusetts Treatment Center (MTC) classification systems to observation cases.
- Jeff Kollath coded risk assessment instruments on a substantial number of cases.
- Yana Litovsky both applied the MTC classification systems to observation cases and coded risk assessment instruments on a substantial number of cases.
- Heather Reu coded risk assessments on an enormous number of cases.
- Elizabeth Schatzel-Murphy edited case files to remove identification information, and scanned the resulting case files, created and managed the Access database, and applied the MTC classification systems to cases.
- Josh Zeier coded risk assessment instruments.

Administrative support and office space was provided for the project by Brandeis University and Sand Ridge Secure Treatment Center.

Table 1

Descriptive Variables for the Bridgewater Observation and Treatment Subsamples

		BOs (n=333)	BTs (n=266)	<i>df</i>	<i>F</i> / χ^2
Age at Release	<i>M</i>	35.54	37.86	1, 560	<i>F</i> = 5.26*
	<i>SD</i>	11.49	12.39		
Race	% Caucasian	90.9%	91.7%	1	χ^2 = .323
Intelligence Score	<i>M</i>	95.09	99.58	1, 451	<i>F</i> = 11.13**
	<i>SD</i>	13.33	15.21		
Number of Non-Sex Victimless Crimes	<i>M</i>	6.44	5.09	1, 485	<i>F</i> = 3.98*
	<i>SD</i>	8.20	6.26		
Number of Non-Sex Victim-Involved	<i>M</i>	1.75	1.50	1, 211	<i>F</i> = 1.34
	<i>SD</i>	1.65	1.38		
Number of Nuisance Sex Crimes	<i>M</i>	2.10	1.35	1, 144	<i>F</i> = 5.63*
	<i>SD</i>	2.12	1.55		
Number of Serious Sex Crimes	<i>M</i>	2.14	2.99	1, 583	<i>F</i> = 22.52**
	<i>SD</i>	1.84	2.49		

* $p < .05$; ** $p < .001$

Table 2

Descriptive Variables for the Three Age-Preference Groups

		Rapists (n=207)	Child Molesters (n=248)	Mixed (n=49)	<i>df</i>	<i>F/χ</i>
Age at Release	<i>M</i>	32.24 _b	38.60 _a	40.01 _a	2, 468	<i>F</i> = 19.05**
	<i>SD</i>	10.22	12.65	11.22		
Race	% Caucasian	86.0%	94.8%	87.8%	2	$\chi^2 = 10.74^*$
Intelligence Score	<i>M</i>	98.74	95.87	100.32	2, 384	<i>F</i> = 2.46
	<i>SD</i>	13.05	38.60	16.93		
Number of Non-Sex Victimless Crimes	<i>M</i>	5.56	6.36	4.31	2, 410	<i>F</i> = 1.38
	<i>SD</i>	6.37	8.90	4.05		
Number of Non-Sex Victim-Involved	<i>M</i>	1.65	1.55	2.07	2, 177	<i>F</i> = .61
	<i>SD</i>	1.55	1.51	2.20		
Number of Nuisance Sex Crimes	<i>M</i>	1.53	1.70	.93	2, 108	<i>F</i> = 1.10
	<i>SD</i>	1.50	2.01	1.14		
Number of Serious Sex Crimes	<i>M</i>	2.10 _b	2.94 _a	3.47 _a	2, 499	<i>F</i> = 12.53**
	<i>SD</i>	1.66	2.61	1.75		

p<.05; ** p<.001

a -- subscript "a" has the highest mean value; groups who do *not* share a subscript were significantly different, $p < .05$, Newman-Keuls.

Table 3

Outline Version of the SRA-Need Assessment Factors

Domain	Factor
Sexual Interests	Child Preference*
	Sexualized Violence*
	Sexual Preoccupation*
Distorted Attitudes	Child Abuse Supportive Beliefs
	Deceitful Women
	Adversarial Sexual Attitudes
	Excessive Sense of Entitlement
	Machiavellianism
Socio-affective functioning	Dysfunctional self-evaluation
	Emotional Congruence with Children*
	Lack of Emotionally Intimate
	Relationships with Adults*
	Callous
	Grievance Thinking*
Self-management	Lifestyle Impulsiveness*
	Dysfunctional Coping*

Table 4

Risk Assessment Instrument Interrater Reliabilities

Scale	Interrater Reliability	N scored by both raters	N scored by at least one rater	<i>M</i>	<i>SD</i>
MnSOST-R	.797	211	560	4.36	5.94
Static-99	.865	232	553	4.56	2.18
Static-2002	.887	258	558	6.58	2.48
RRASOR	.886	297	567	2.33	1.29
RM2000/S	.821	281	566	2.65	0.92
ASOAP	.714	223	542	0.46	0.14
SORAG	.881	105	537	8.70	12.39
SVR-20	.602	116	540	14.36	5.92
SRA Need	.713	89	513	15.24	5.81
PCL-R	.807	115	555	13.67	6.50

Table 5

AUC Values for Serious Sexual Charges for 3, 10, and 15 Years for the Entire MTC Sample

Variable	3 Years		10 Years		15 Years	
	AUC	CI	AUC	CI	AUC	CI
PCL-R Factor 1	.567	.485-.648	.569	.498-.641	.604*	.508-.700
PCL-R Factor 2	.630**	.551-.709	.632***	.562-.703	.634*	.536-.733
PCL-R Total	.650***	.573-.726	.641***	.572-.710	.657**	.565-.749
Mn-SOST-R	.684***	.618-.749	.672***	.603-.742	.664**	.564-.765
RM2000 Sexual Risk	.674***	.603-.745	.644***	.575-.714	.633**	.538-.727
RM2000 Violence Risk	.574	.488-.661	.539	.462-.616	.548	.446-.650
RM2000 Combined Risk	.654***	.576-.732	.610**	.536-.683	.600	.502-.698
Static 99	.713***	.650-.777	.684***	.619-.749	.647**	.557-.736
Doren	.703***	.642-.765	.694***	.630-.759	.658**	.570-.746
RRASOR	.669***	.603-.735	.681***	.615-.748	.649**	.559-.739
Static 2002	.680***	.611-.749	.674***	.608-.740	.671***	.581-.762
A-SOAP Sexual Part 1	.652***	.575-.730	.684***	.613-.754	.643**	.546-.739
A-SOAP Impulsivity Part 2	.649***	.568-.730	.622**	.547-.698	.608*	.504-.711
A-SOAP Intervention Part 3	.490	.401-.580	.496	.419-.573	.566	.469-.663
A-SOAP Static total	.691***	.619-.763	.694***	.626-.762	.665**	.566-.764
A-SOAP Dynamic total	.494	.406-.583	.496	.419-.572	.563	.467-.659
A-SOAP Total	.646***	.572-.721	.652***	.582-.721	.655**	.559-.751
VRAG Total score	.645***	.564-.725	.631***	.559-.703	.606	.505-.707
SORAG Total score	.671***	.591-.752	.673***	.603-.742	.638**	.540-.736
SVR20	.655***	.583-.726	.684***	.620-.749	.677***	.588-.766
SRA sexualization	.630**	.547-.713	.672***	.599-.746	.686***	.590-.782
SRA social affective	.682***	.612-.751	.696***	.631-.762	.694***	.598-.790
SRA self management	.627**	.550-.703	.644***	.573-.715	.683***	.590-.777
SRA Need Assessment	.702***	.637-.768	.734***	.672-.795	.756***	.669-.843
SRA 7	.648***	.576-.721	.682***	.615-.750	.676***	.582-.771
APD DSM IV criteria	.556	.464-.648	.616**	.531-.701	.631*	.511-.751
Conduct Disorder DSM IV criteria	.639**	.548-.730	.649***	.564-.733	.587	.465-.708
<i>M</i>	.640		.643		.641	
<i>SD</i>	.057		.059		.046	

* p < .05 ** p < .01 *** p < .001

Table 6

Mean AUC Values for the Entire MTC Sample and All Subsamples for Different Crime Types for 3, 10, and 15 Years

Sample	Crime Type	3 Years		10 Years		15 Years	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Entire Sample	Serious Sexual	.640	.057	.643	.059	.641	.046
	Non-Sexual Victim-Involved	.601	.060	.613	.064	.630	.057
	Non-Sexual Victimless	.579	.067	.594	.058	.626	.063
Bridgewater Treatment	Serious Sexual	.569	.061	.580	.062	.556	.050
	Non-Sexual Victim-Involved	.552	.075	.591	.089	.638	.081
	Non-Sexual Victimless	.568	.079	.588	.085	.601	.080
Bridgewater Observation	Serious Sexual	.624	.069	.611	.057	.630	.050
	Non-Sexual Victim-Involved	.608	.061	.604	.061	.614	.064
	Non-Sexual Victimless	.626	.064	.596	.048	.625	.068
Rapists	Serious Sexual	.657	.040	.625	.070	.629	.090
	Non-Sexual Victim-Involved	.629	.058	.636	.071	.644	.068
	Non-Sexual Victimless	.604	.084	.599	.064	.608	.087
Child Molesters	Serious Sexual	.596	.072	.632	.078	.633	.071
	Non-Sexual Victim-Involved	.548	.059	.585	.058	.601	.068
	Non-Sexual Victimless	.573	.064	.578	.055	.616	.062

Table 7

Summary of t-tests Comparing Mean AUC Values for Serious Sexual Charges to Non-sexual Victim-Involved and Victimless Charges for All predictive Measures

Sample	Crime Comparison	3 Years			10 Years			15 Years		
		<i>t</i>	<i>p</i>	<i>dir</i>	<i>t</i>	<i>p</i>	<i>dir</i>	<i>t</i>	<i>p</i>	<i>dir</i>
Entire Sample	Serious Sex vs. Victim-Involved	2.78	< .01	S>VI	1.76			.72		
	Serious Sex vs. Victimless	2.58	< .05	S>V	3.16	< .01	S>V	1.00		
Rapists	Serious Sex vs. Victim-Involved	2.54	< .05	S>VI	-.685			-.80		
	Serious vs. Victimless	3.14	< .01	S>V	1.33			.79		
Child Molesters	Serious Sex vs. Victim-Involved	2.39	< .05	S>VI	2.01			1.42		
	Serious Sex vs. Victimless	1.04			2.74	< .05	S>V	.967		

S = serious sexual charges

VI = victim-involved non-sexual charges

V = victimless, non-sexual charges

Table 8

Summary of t-tests Comparing Mean AUC Values for Serious Sexual Charges to Non-sexual Victim-Involved and Victimless Charges for Sexual Offender Specific Predictive Measures

Sample	Crime Comparison	3 Years			10 Years			15 Years		
		<i>t</i>	<i>p</i>	<i>dir</i>	<i>t</i>	<i>p</i>	<i>dir</i>	<i>t</i>	<i>p</i>	<i>dir</i>
Entire Sample	Serious Sex vs. Victim-Involved	4.08	< .001	S>VI	3.25	< .01	S>VI	2.18	< .05	S>VI
	Serious Sex vs. Victimless	3.25	< .001	S>V	4.69	< .001	S>V	2.75	< .05	S>V
Rapists	Serious Sex vs. Victim-Involved	4.29	< .001	S>VI	.850			.306		
	Serious vs. Victimless	5.95	< .001	S>V	2.18	< .05	S>V	1.36		
Child Molesters	Serious Sex vs. Victim-Involved	2.90	< .01	S>VI	2.67	< .05	S>VI	2.31	< .05	S>VI
	Serious Sex vs. Victimless	1.89			4.10	< .001	S>V	3.04	< .01	S>V

S = serious sexual charges

VI = victim-involved non-sexual charges

V = victimless, non-sexual charges

Table 9

Summary of t-tests Comparing Mean AUC Values for Serious Sexual Charges to Non-sexual Victim-Involved and Victimless Charges for General Criminal Predictive Measure

Sample	Crime Comparison	3 Years			10 Years			15 Years		
		<i>t</i>	<i>p</i>	<i>dir</i>	<i>t</i>	<i>p</i>	<i>dir</i>	<i>t</i>	<i>p</i>	<i>dir</i>
Entire Sample	Serious Sex vs. Victim-Involved	-0.82			-2.04			-2.38	< .05	S<VI
	Serious Sex vs. Victimless	-0.86			1.76			-2.34	< .05	S<V
Rapists	Serious Sex vs. Victim-Involved	-0.97			-3.23	< .05	S<VI	-3.10	< .05	S<VI
	Serious Sex vs. Victimless	-0.22			-0.64			-0.77		
Child Molesters	Serious Sex vs. Victim-Involved	-0.26			-0.49			-0.58		
	Serious vs. Victimless	-2.51			-0.93			-2.34	< .05	S<V

S = serious sexual charges

VI = victim-involved non-sexual charges

V = victimless, non-sexual charges

Table 10

For Serious Sexual Charges the Number of Comparisons in Which Each Predictive Measure Had a Significantly Lower AUC than Another Measure at Each Significance Level for the 3 and 10-Year Gates

Predictive Measure	3 Years			10 Years		
	$p < .05$	$p < .01$	$p < .001$	$p < .05$	$p < .01$	$p < .001$
PCL-R Factor 1	4	3	0	9	3	1
RM2000 Violence Risk	5	2	0	5	6	7
A-SOAP Intervention Part 3	0	4	17	1	4	19
A-SOAP Dynamic total	0	7	13	1	4	19
DSM IV APD	6	3	1	0	0	1

Table 11

For Victim-Involved, Non-Sexual Charges the Number of Comparisons in Which Each Predictive Measure Had a Significantly Higher (+) or Lower AUC than Another Measure at Each Significance Level for the 3 and 10-Year Gates

Predictive Measure	3 Years			10 Years		
	<i>p</i> < .05	<i>p</i> < .01	<i>p</i> < .001	<i>p</i> < .05	<i>p</i> < .01	<i>p</i> < .001
RM2000 Violence Risk	6(+)	1(+)	9(+)	1(+)	4(+)	7(+)
RM2000 Combined Risk	1(+)	2(+)	7(+)	0	3(+)	4(+)
Mn-SOST-R	2(+)	2(+)	4(+)	1(+)	0	6(+)
A-SOAP Sexual Part 2	2	7	8	2	5	12
SRA sexualization	1	5	12	0	3	17
Doren	2	1	1	3	3	2
RRASOR	2	4	4	0	2	8

Table 12

For Victimless Charges the Number of Comparisons in Which Each Predictive Measure Had a Significantly Lower AUC than Another Measure at Each Significance Level for the 3 and 10-Year Gates

Predictive Measure	3 Years			10 Years		
	$p < .05$	$p < .01$	$p < .001$	$p < .05$	$p < .01$	$p < .001$
PCL-R Factor 1	2	2	7	4	1	6
RM2000 Sexual Risk	2	2	3	2	1	1
Doren	1	5	3	3	4	5
RRASOR	2	2	8	2	4	7
A-SOAP Sexual, Part 1	2	3	17	0	4	10
A-SOAP Intervention, Part 3	2	5	3	4	2	9
SRA sexualization	1	1	22	1	1	16
SRA social affective	1	2	7	1	4	1
SRA 7	3	5	9	2	2	5
DSM IV Conduct Disorder	7	2	9	2	3	6

Table 13

Eigenvalues and Percent of Variance Explained for the Principle Components Analysis of Actuarial Scale Items

Factor	Eigenvalue	Percent of Variance
Criminal Persistence	14.00	17.3%
Sexual Persistence	9.50	11.7%
Young & Single	6.30	7.8%
Violent Sexual Assaults	4.91	6.1%
Male Victim Choice	4.35	5.4%

Table 14

Varimax Rotated Component Matrix for the Principle Components Analysis of Actuarial Scale Items

	Component				
	1	2	3	4	5
VRAG nonviolent *	.730				
RM Criminal appearances	.726	.242			
SVR Past nonsexual nonviolent offenses	.720				
S02 Arrest/Sentencing occasions	.719	.348			
S99 Prior sentencing dates	.712	.243			
ASO Multiple types of offenses	.707			.244	
ASO Adult antisocial behavior	.686	.200			
VRAG Failure on prior conditional release	.672	.290			
S02 Any breach of conditional release	.671	.294			
ASO Juvenile antisocial behavior	.671		.294		
SVR Past supervision failure	.655	.268			
SVR Past nonsexual violent offense	.622			.208	
RM Violent appearances	.616			.393	
MNS adolescent antisocial	.614		.282		
ASO Ever charged before age 16	.593		.290		
RM Burglary	.593				
S99 Prior non-sexual violence	.589				
S02 Prior non-sexual violence	.579				
S02 Years free prior to index offense	.510	.320	.200		
MNS Under supervision	.481	.302			
VRAG Elementary school maladjustment	.412		.260		.210
SVR Employment problems	.373		.242		
SVR Substance use problems	.366			.257	
VRAG History of alcohol problems	.352			.265	
VRAG Lived biological parents to 16	.341				
MNS drug or alcohol abuse	.325				
MNS Employment history	.303		.269		
ASO Physical abuse/vicarious violence history					
S02 Prior sexual App		.849			
RM Sexual appearances	.221	.843			.206
S99 Prior sex offenses	.217	.815			
ASO Prior legally charged sex off	.265	.809			
MNS No. sex-related convictions		.712			
S02 High rate sexual offending	.202	.712			
SOR No. previous sex offense convictions	.261	.684			.250
ASO No. of sex abuse victims		.609			.378

S02 Non-contact sex offenses		.608			-.311
RM Non-contact sex offense		.605			-.323

Table 14 (cont.)

S99 non-contact sex offense convictions		.603			-.311
ASO Duration sex offense history		.550	-.280	-.268	
SOR Criminal history score violent off	.443	.496			
SVR High density sex offense		.457			.292
MNS No. of diff age groups sex-related off		.454			.328
SVR Multiple sex offense types		.423			
MNS sex offense in a public place		.385		.312	
S02 2 or more victims > 12, 1 unrelated		.360		-.237	.300
S02 Juvenile arrest sexual off	.250	.321	.267		
MNS Length of sex offending history		.263			
RM Age Sex			.880		
RM Age Violence			.864		
S02 Age at release			.862		
VRAG Age index offense			.769		
MNS Age at release			.767		
S99 Young			.657		
S99 Lived with lover 2 years			.639		.368
RM Single			.596		.369
VRAG Never married			.583		.401
MNS Mult. sex acts on single victim in 1 contact					
S02 Stranger victims		.407		.770	
RM Stranger victim of sex offense		.413		.764	
S99 Any stranger victims		.413		.763	
MNS Sex victim a stranger		.426		.751	
ASO Expressive aggression in sex off	.280	-.212		.685	
SVR Phys harm to sex victim	.238	-.235		.646	
MNS Force used sex off	.266			.619	
SVR Used weapons/ death threats sex off	.215	-.201		.606	
S99 Non-sexual violence	.242			.578	
VRAG Victim injury low		.281		-.540	
SOR Sex offenses only girls under 14				.444	.392
S99 Any unrelated victims?		.285	.280	.414	.245
S02 Any unrelated victims?		.283	.254	.406	.239
SVR Escalation severity sex off		.307		.356	
RM Male victim of sex off					.871
S99 Male victims					.869
S02 Male victims					.867
ASO Male child victim					.830
VRAG No female victim					.781
SVR Victim of sexual abuse					.321
ASO Sexual victimization history					.318

MNS Offended against 13- to 15-year					.313
ASO Degree of planning in sex off					.278

* ASO -- The Adult Sex Offender Assessment Protocol (A-SOAP-II); MNS -- The Minnesota Sex Offender Screening Tool (Mn-SOST-R);
 RM -- Risk Matrix 2000; S02 -- Static-2002; S99 -- Static-99; SVR -- Sexual Violence Risk-20; VRAG -- Violence Risk Assessment Guide

Table 15

AUCs and Confidence Intervals for the Five Factors Derived from the PCA of Actuarial Items

Factor	5 years		10 years	
	<i>AUC</i>	<i>CI</i>	<i>AUC</i>	<i>CI</i>
Criminal Persistence	.586*	.509-.663	.599*	.520-.677
Sexual Persistence	.661***	.590-.731	.668***	.593-.742
Young & Single	.494	.420-.569	.485	.408-.561
Violent Sexual Assaults	.585*	.511-.658	.595*	.519-.670
Male Victim Choice	.654***	.581-.728	.629**	.552-.705

* p < .05 ** p < .01 *** p < .001

Table 16

Correlations between Established Actuarial Instruments and the Five Static Historical Factors

	MnSOST-R	Static-99	Static-2002	RRASOR	RM2000/S	SORAG
Criminal Persistence	.38	.43	.40	.16	.77	.76
Sexual Persistence	.53	.63	.72	.71	.62	.26
Young & Single	.22	.21	.37	.13	.42	.29
Violent Stranger	.40	.40	.18	.00	.21	.19
Male Victim Choice	.14	.26	.18	.48	.24	.24

Table 17

Percent of Variance in Established Actuarial Scales Accounted for by the Five Static Historical Factors

Scale	% Variance Accounted for
MnSOST-R	66%
Static-99	85%
Static-2002	88%
RRASOR	78%
RM2000/S	73%
SORAG	83%

Table 18

Logistic Regressions Using Static Historical and Need Factors to Predict Serious Sexual Recidivism at 5- (n=346) and 10- (n = 282) Year Follow-up Outcomes

	Predictors	B	(SE)	p	E (B)
5-Year	Composite of static historical factors	.564	(.187)	.003	1.757
Gate	SRA-Need Assessment	.542	(.171)	.001	1.720
	Constant	-1.763			
10-Year	Composite of static historical factors	.541	(.189)	.004	1.718
Gate	SRA-Need Assessment	.655	(.178)	.000	1.924
	Constant	-1.427			

Table 19

Predictive Accuracy of the Combination of the Static Historical Composite and the SRA Need Scale in Predicting Serious Sexual Recidivism at 5- and 10-Year Follow-up Outcomes

Predictors	5-year		10-year	
	<i>AUC</i>	<i>CI</i>	<i>AUC</i>	<i>CI</i>
Composite of Static Historical Variables	.710	.647-.773	.717	.650-.784
SRA - Need Assessment	.729	.663-.795	.747	.681-.813
Combination	.756	.697-.815	.772	.710-.834

Table 20

Logistic Regressions Using Age at Index Offense and Age at Discharge to Predict Serious Sexual Recidivism at 5- and 10- Year Follow-up Outcomes

	Predictors	b	p
5-Year Gate	Youthful Index Offense	.244	.013
	Age on Discharge	.256	.007
	Age on Discharge Squared	-0.003	.016
	Constant	-6.792	
10-Year Gate	Youthful Index Offense	.219	.014
	Age on Discharge	.197	.022
	Age on Discharge Squared	-.002	.050
	Constant	-5.794	

Table 21

Serious Sexual Recidivism by Age at Index and Age at Discharge

Age at Index Offense	Age at Discharge	Sexual Recidivism	
		5-Year	10-Year
27 & Under	30 & under	16%	16%
	Over 30 to 45	33%	44%
	Over 45 to 59		
	60+		
28 - 38	30 & under	14%	14%
	Over 30 to 45	20%	15%
	Over 45 to 59	42%	36%
	60+	0%	0%
Older	30 & under		
	Over 30 to 45	10%	10%
	Over 45 to 59	19%	16%
	60+	10%	9%

Table 22

Discriminant Function Analyses Predicting Serious Sexual Recidivism for Child Molesters

Time Gate	Independent Variables	Discrim. Func. Coefficient	Structural Coefficient	Canonical <i>r</i>	χ^2	<i>df</i>	<i>p</i>
3-Year	Paraphilias	.668	.794	.302	15.619	2	< .001
	High fixation	.620	.757				
	Social isolation		.380				
	Adult sex crimes		.320				
	Male Victim		.249				
	Impulsivity		-.009				
5-Year	Paraphilias	.536	.691	.435	32.097	4	< .001
	Social isolation	.366	.618				
	Impulsivity	.455	.523				
	Adult sex crimes	.448	.371				
	High fixation		.326				
	Male Victim		.248				
7-Year	Paraphilias	.585	.663	.493	39.444	3	< .001
	Adult sex crimes	.630	.604				
	Impulsivity	.563	.411				
	Social isolation		.291				
	High fixation		.243				
	Male Victim		.203				
10-Year	Social isolation	.437	.632	.500	35.397	4	< .001
	Adult sex crimes	.556	.619				
	Paraphilias	.375	.595				
	Impulsivity	.431	.362				
	High fixation		.327				
	Male Victim		.227				

Table 22 (*cont.*)

15-Year	Social isolation	1.000	1.000	.413	12.969	1	< .001
	High fixation		.306				
	Paraphilias		.256				
	Adult sex crimes		.218				
	Male Victim		.216				
	Impulsivity		.050				

Table 23

Discriminant Function Analyses Predicting Serious Sexual Recidivism for Rapists for the Entire Sample

Time Gate	Independent Variables	Discrim. Func. Coefficient	Structural Coefficient	Canonical <i>r</i>	χ^2	<i>df</i>	<i>p</i>
3-Year	Pervasive Anger	0.856	0.857	0.320	15.797	2	< .001
	Offense Planning	0.516	0.516				
	Juv unsoc. agg.		0.373				
	Adult unsoc. agg.		0.332				
	Sadism		0.045				
	Adult sex crimes		0.04				
5-Year	Pervasive Anger	0.736	0.743	0.317	14.311	2	< .001
	Offense Planning	0.669	0.667				
	Juv unsoc. agg.		0.339				
	Adult unsoc. agg.		0.261				
	Sadism		0.102				
	Adult sex crimes		0.079				
7-Year	Offense Planning	1.000	1.000	0.266	9.064	1	< .005
	Pervasive Anger		0.019				
	Juv unsoc. agg.		-0.004				
	Adult unsoc. agg.		-0.153				
	Sadism		0.208				
	Adult sex crimes		0.254				
10-Year	Offense Planning	1.000	1.000	0.255	6.579	1	< .01
	Pervasive Anger		0.025				
	Juv unsoc. agg.		0.002				
	Adult unsoc. agg.		-0.148				
	Sadism		0.271				
	Adult sex crimes		0.241				

Table 23 (cont.)

15-Year	Offense Planning Pervasive Anger Juv unsoc. agg. Adult unsoc. agg. Sadism Adult sex crimes	**	**	**	**	= .051
---------	---	----	----	----	----	--------

** no variable reached the .05 criterion for selection.

Figure Captions

Figure 1. A histogram of the number of offenders who fell into each ten-year age interval for the age of their first offense.

Figure 2. A histogram of the number of offenders who fell into each ten-year age interval for their age at the offense that led to their observation at MTC.

Figure 3. A graph of the survival curves depicting the time to the first serious sexual charge for the BO and the BT subsamples.

Figure 4. A graph of the survival curves depicting the time to the first victim-involved, non-sexual charge for the BO and the BT subsamples.

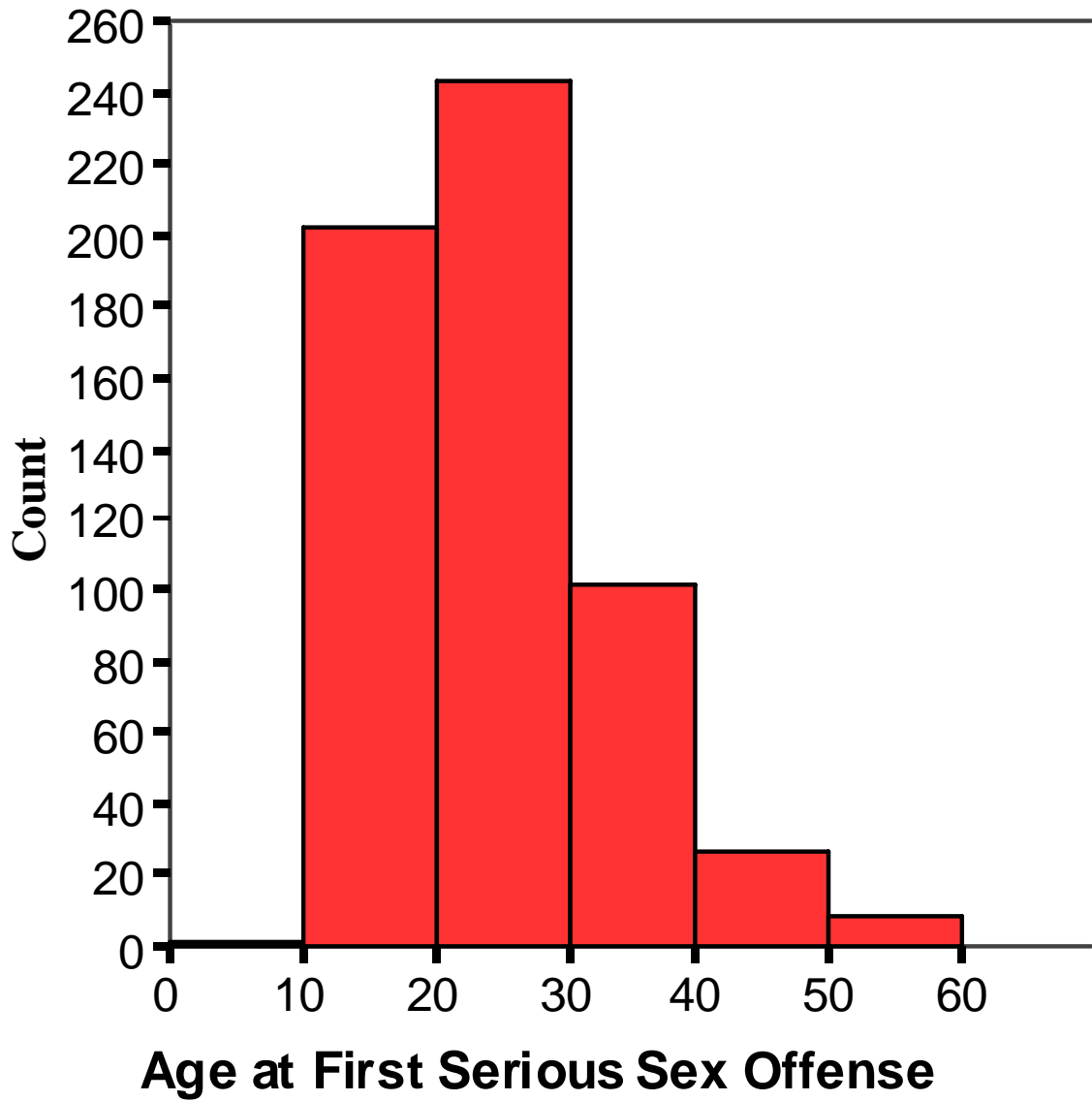
Figure 5. A graph of the survival curves depicting the time to the first victimless, non-sexual charge for the BO and the BT subsamples.

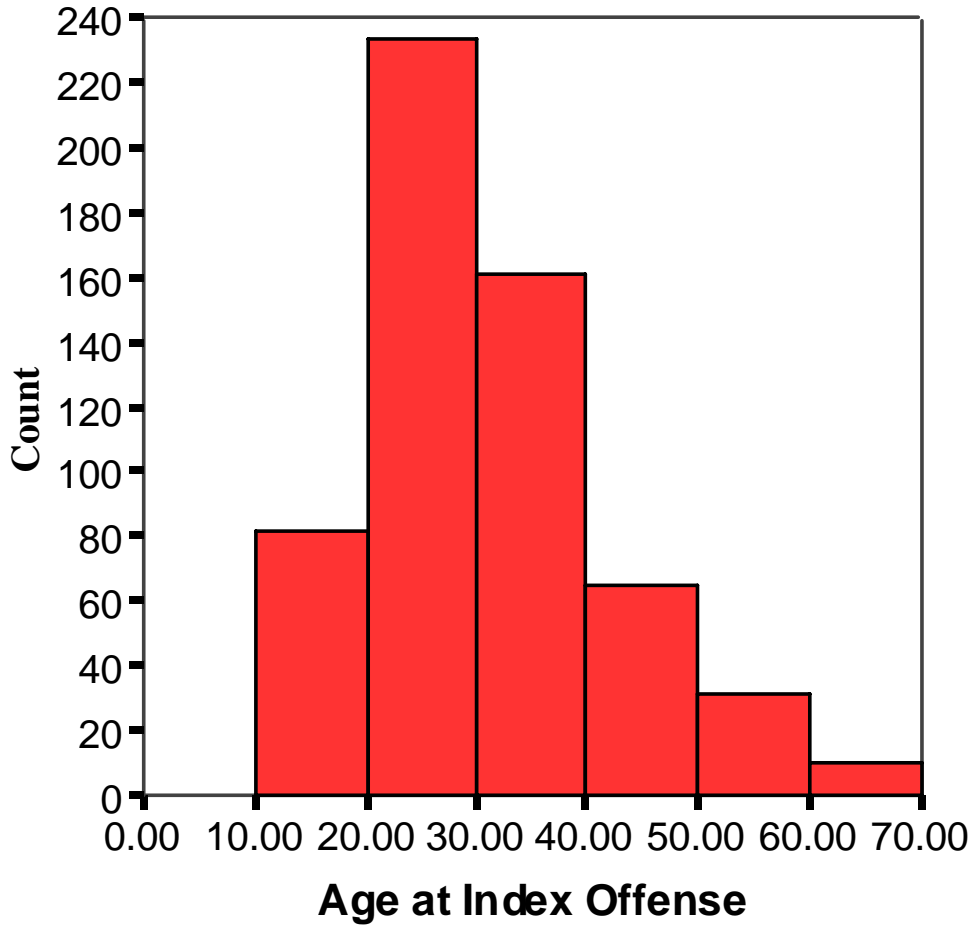
Figure 6. A graph of the survival curves depicting the time to first victim-involved, non-sexual charge for the Random BO and the BT subsamples.

Figure 7. A graph of the survival curves depicting the time to first serious sexual charge for the child molester and the rapist subsamples.

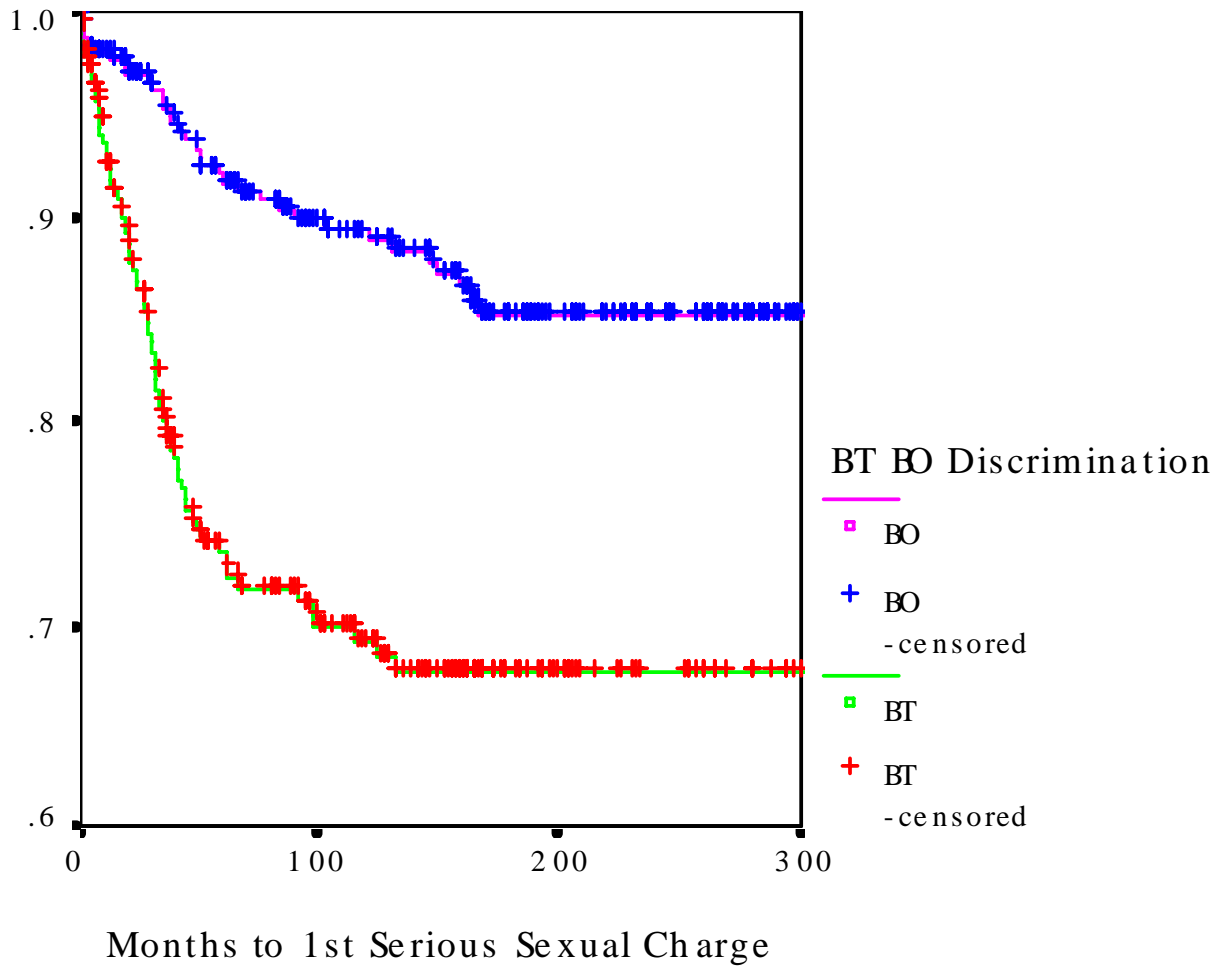
Figure 8. A graph of the survival curves depicting the time to first victim-involved, non-sexual charge for the child molester and the rapist subsamples.

Figure 9. A graph of the survival curves depicting the time to first victimless, non-sexual charge for the child molester and the rapist subsamples.

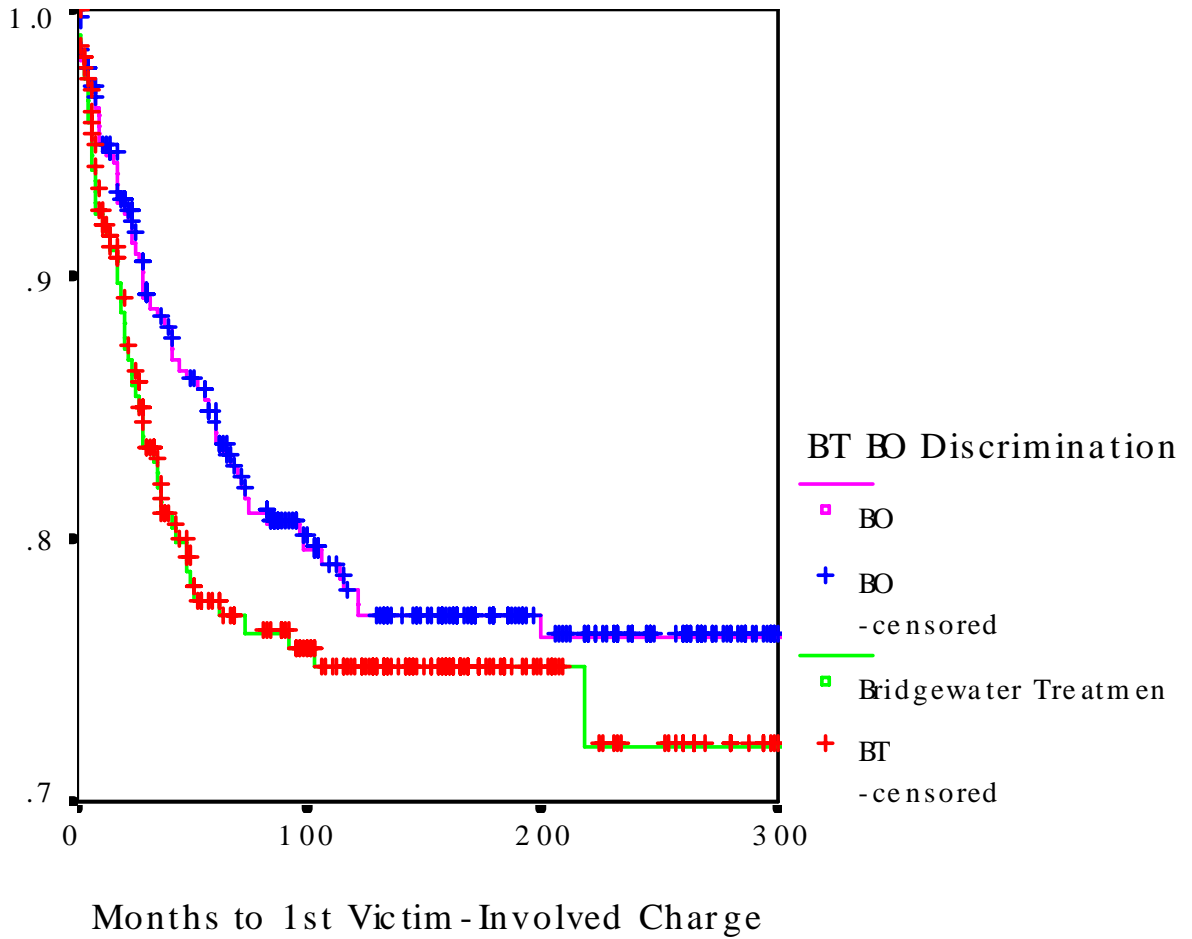




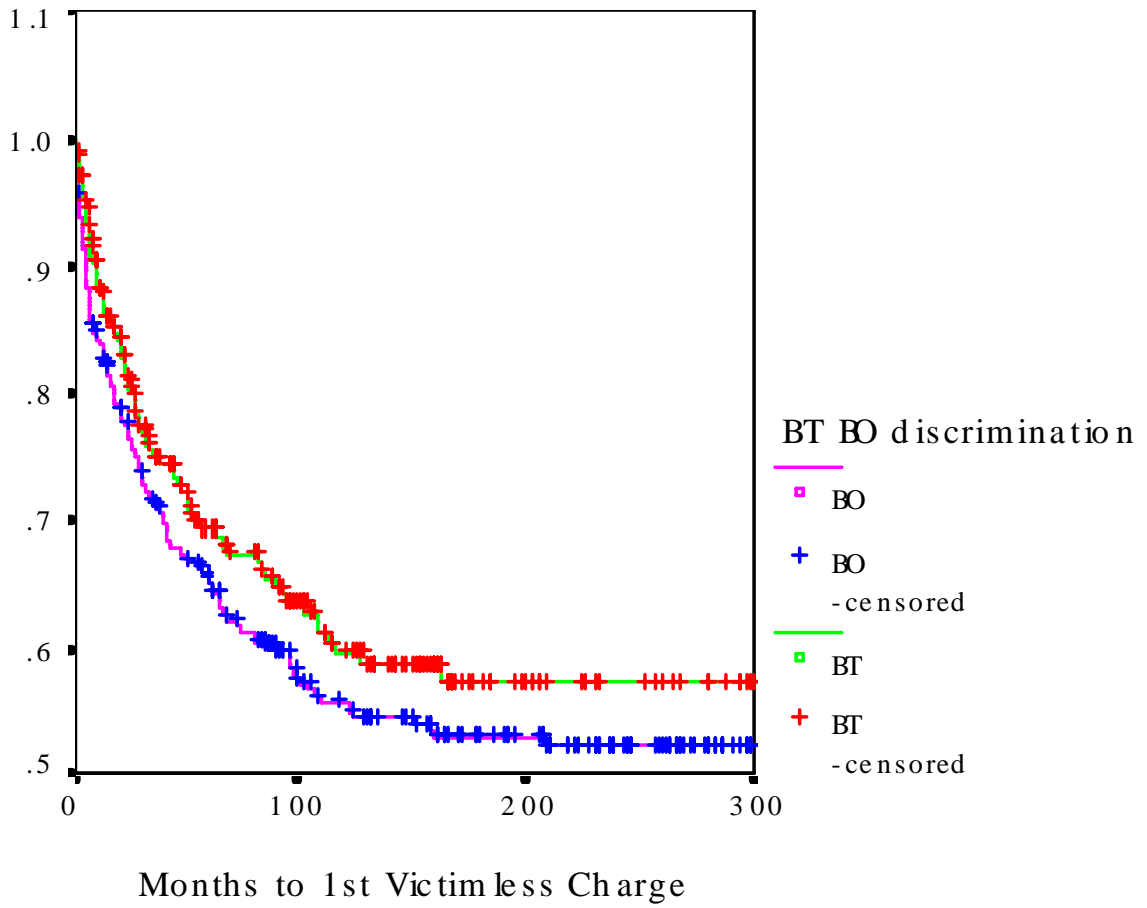
BO BT Survival: Serious Sex Charge



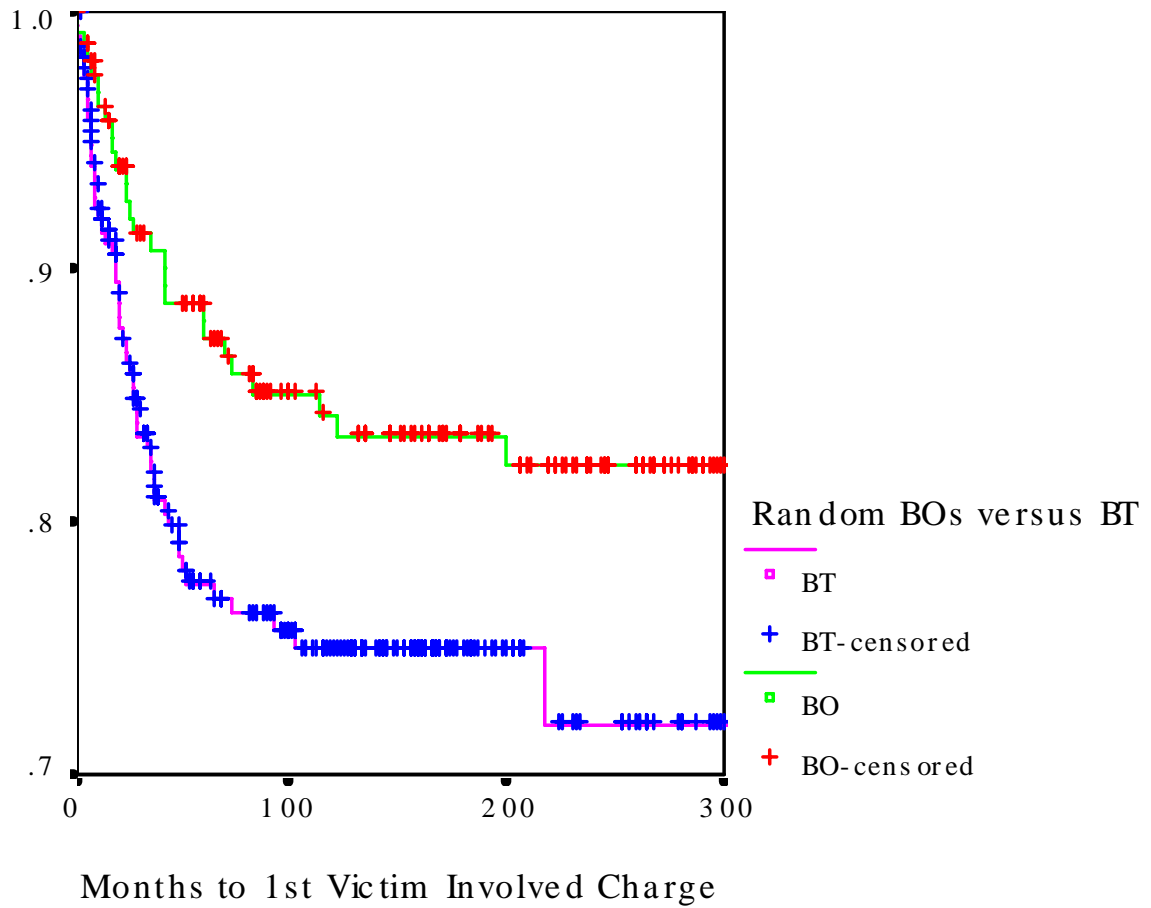
BO BT Survival: Victim - Involved



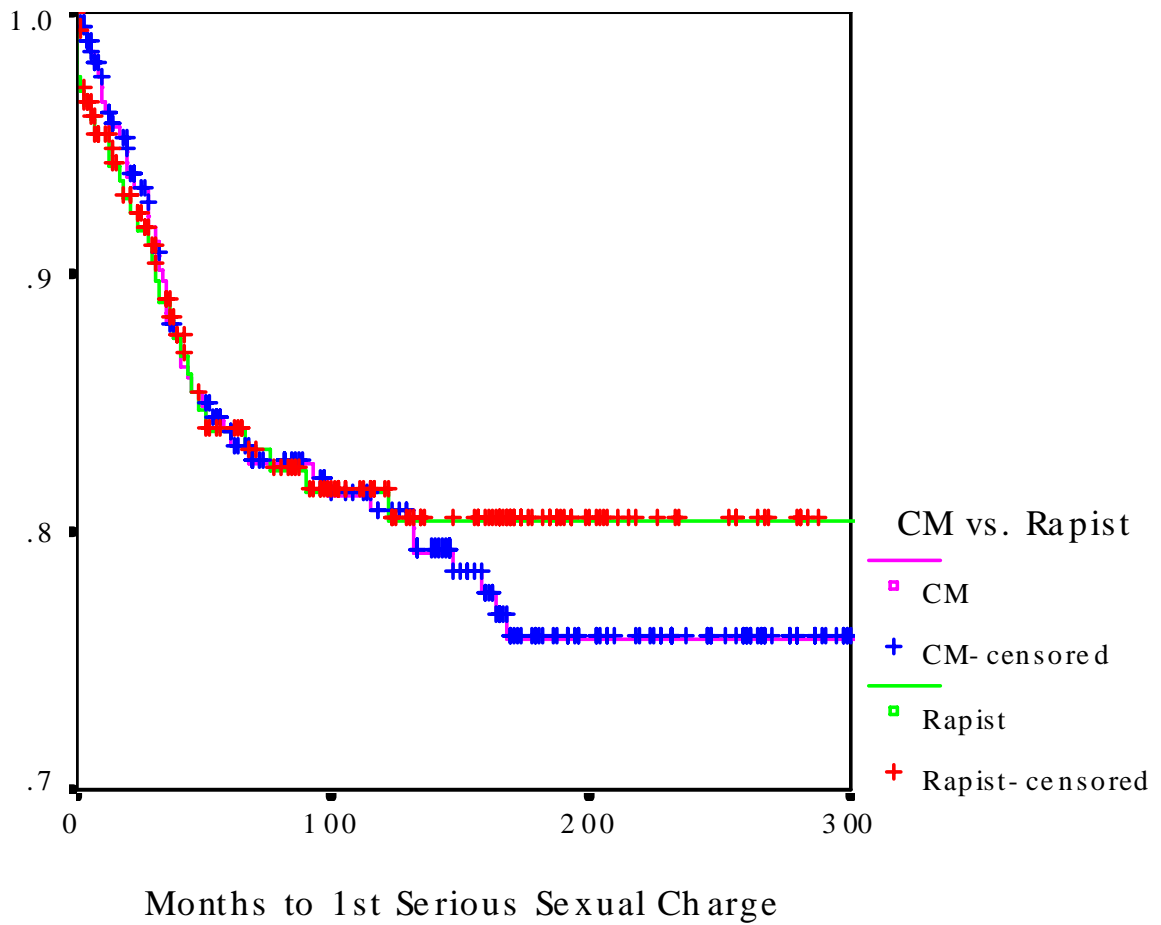
BT/ BO Survival: Victimless Charge



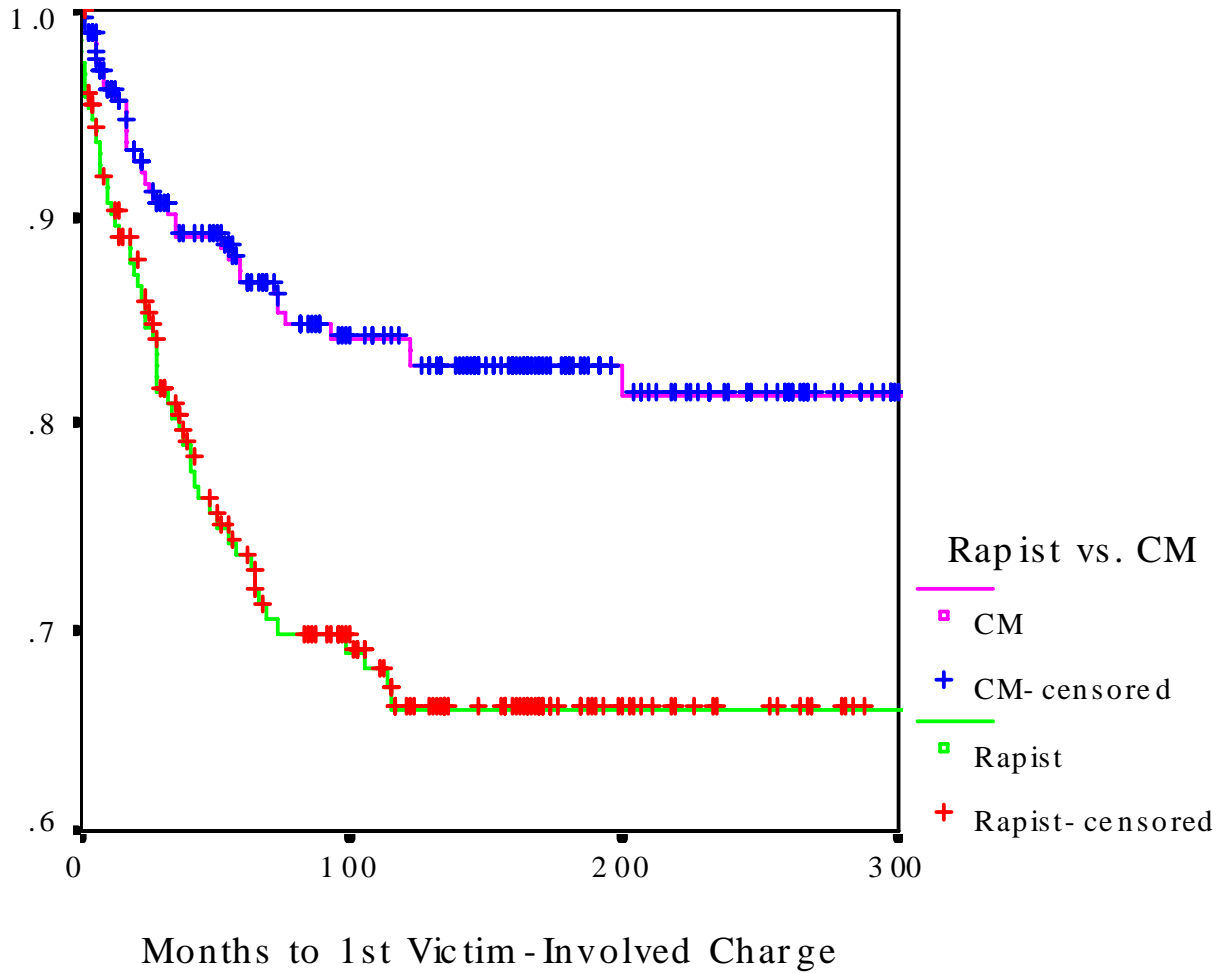
Random BO/ BT Survival: Vic- Involve



Rapist CM Survival: Serious Sex Charge



Rapist CM Survival: Vic- Involved Ch



Rapist CM Survival: Victimless Ch

