Incremental Validity of the Static-99 and the Personality Assessment Inventory for Predicting Sexual Recidivism Among Adult Male Sex Offenders

Christopher Scott Brown
Pacific University

Recommended Citation
Incremental Validity of the Static-99 and the Personality Assessment Inventory for Predicting Sexual Recidivism Among Adult Male Sex Offenders

Abstract
The Static-99 is an established risk assessment instrument designed for use with adult male sex offenders. This instrument demonstrates significant utility, but because of the importance of making accurate decisions regarding the dangerousness of sex offenders, improving the accuracy of the instrument is critical. The current study examined whether the addition of data from the Antisocial Features and Aggression scales of the Personality Assessment Inventory (PAI) would improve the predictive power of the Static-99 in a sample of 72 adult male sex offenders evaluated at a community clinic. Neither the Static-99 nor any of the modifications incorporating the PAI scales significantly predicted sexual rearrest or sexual reconviction during the follow-up period, which averaged 5.2 years. Implications of these findings are discussed in the context of the sample characteristics.

Degree Type
Thesis

Degree Name
Master of Science in Psychology

Committee Chair
Michelle R. Guyton, Ph.D.

Second Advisor
Genevieve Arnaut, Psy.D., Ph.D.

Subject Categories
Psychiatry and Psychology

This thesis is available at CommonKnowledge: https://commons.pacificu.edu/spp/3
Copyright and terms of use

If you have downloaded this document directly from the web or from CommonKnowledge, see the “Rights” section on the previous page for the terms of use.

If you have received this document through an interlibrary loan/document delivery service, the following terms of use apply:

Copyright in this work is held by the author(s). You may download or print any portion of this document for personal use only, or for any use that is allowed by fair use (Title 17, §107 U.S.C.). Except for personal or fair use, you or your borrowing library may not reproduce, remix, republish, post, transmit, or distribute this document, or any portion thereof, without the permission of the copyright owner. [Note: If this document is licensed under a Creative Commons license (see “Rights” on the previous page) which allows broader usage rights, your use is governed by the terms of that license.]

Inquiries regarding further use of these materials should be addressed to: CommonKnowledge Rights, Pacific University Library, 2043 College Way, Forest Grove, OR 97116, (503) 352-7209. Email inquiries may be directed to: copyright@pacificu.edu
ABSTRACT

The Static-99 is an established risk assessment instrument designed for use with adult male sex offenders. This instrument demonstrates significant utility, but because of the importance of making accurate decisions regarding the dangerousness of sex offenders, improving the accuracy of the instrument is critical. The current study examined whether the addition of data from the Antisocial Features and Aggression scales of the Personality Assessment Inventory (PAI) would improve the predictive power of the Static-99 in a sample of 72 adult male sex offenders evaluated at a community clinic. Neither the Static-99 nor any of the modifications incorporating the PAI scales significantly predicted sexual rearrest or sexual reconviction during the follow-up period, which averaged 5.2 years. Implications of these findings are discussed in the context of the sample characteristics.
ACKNOWLEDGMENTS

A team of graduate-level researchers gathered the data used in this study. This group was supervised by Michelle R. Guyton, Ph.D., coordinated by Erica Vo, M.S., and included Kathryn Marshall, M.S., Jacqueline Means, M.S., Pamela Buchanan, M.S., and me. I greatly appreciate the assistance of these colleagues. I would also like to thank my family and my close friends, who provided me with a year of priceless support and encouragement.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>3</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>5</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>6</td>
</tr>
<tr>
<td>Sex Offender Recidivism</td>
<td>6</td>
</tr>
<tr>
<td>Actuarial Prediction of Sex Offender Recidivism</td>
<td>14</td>
</tr>
<tr>
<td>Static-99</td>
<td>16</td>
</tr>
<tr>
<td>Personality Assessment Inventory</td>
<td>23</td>
</tr>
<tr>
<td>Rationale for Current Study</td>
<td>26</td>
</tr>
<tr>
<td>METHOD</td>
<td>28</td>
</tr>
<tr>
<td>Sample Characteristics</td>
<td>28</td>
</tr>
<tr>
<td>Instruments</td>
<td>30</td>
</tr>
<tr>
<td>Procedure</td>
<td>31</td>
</tr>
<tr>
<td>RESULTS</td>
<td>33</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>37</td>
</tr>
<tr>
<td>Review of Findings</td>
<td>37</td>
</tr>
<tr>
<td>Implications of Findings</td>
<td>37</td>
</tr>
<tr>
<td>Strengths and Limitations of Current Study</td>
<td>39</td>
</tr>
<tr>
<td>Directions of Future Study</td>
<td>40</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>42</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Static-99 Items and Scoring ................................................................. 17
2. Descriptive Statistics ........................................................................... 29
3. Numbers of Subjects in Various Risk Categories across ANT and AGG Levels ...... 34
4. Correlations between Independent Variables ........................................... 34
5. Static-99 Scores across ANT and AGG Groupings .................................. 35
6. Binary Logistic Regressions ................................................................. 36
7. Areas under ROC Curves ..................................................................... 37
INTRODUCTION

Sex Offender Recidivism

According to the Bureau of Justice Statistics, “Persons age 12 or older experienced an average annual 140,990 completed rapes, 109,230 attempted rapes, and 152,680 completed and attempted sexual assaults between 1992 and 2000” (Rennison, 2002, p. 1). Such sexual assault has serious physical and psychological consequences for victims (Foa & Steketee, 1987). Because of their prevalence and impact upon victims, sex offenses have been the subject of an enormous amount of empirical research. In this study, I attempted to add to that body of knowledge by improving the predictive accuracy of one psychological assessment instrument that is commonly used with sex offenders. I first examine the recidivism rates of sex offenders and actuarial tools that are often used to predict such recidivism. I present the results of an analysis attempting to improve upon a widely used actuarial risk assessment instrument by adding data on antisocial and psychopathic traits derived from a personality inventory.

Recidivism Rates

Researchers have examined the recidivism rates of sex offenders extensively. They have frequently broken down recidivism across original offense type (e.g., child molestation, rape, other) and type of recidivism (e.g., general, nonsexual violent, sexual). The level of governmental organization associated with prosecution, incarceration, probation, and parole make tracking recidivism relatively easy and accurate as long as the offense is reported and prosecuted. However, consumers of this research must remember
that the samples used in all sex offense research represent the minority of sex offenders who have been charged with crimes (Rennison, 2002). For example, Bureau of Justice Statistics data in Rennison’s report indicate that between 1992 and 2000, only 35% of attempted rapes, 37% of completed rapes, and 26% of completed and attempted sexual assaults against females were reported to the police, let alone successfully prosecuted. Prevalence rates are similarly underestimated in recidivism studies, because researchers can only analyze recidivism patterns if an original crime and subsequent crimes were reported and, if the inclusion criterion is conviction, successfully prosecuted.

Hanson and Morton-Bourgon (2005) conducted one of the largest meta-analyses of sex offender recidivism to date, examining 29,450 offenders in 82 studies. The authors performed their analyses using $d$ as an alternative to correlation coefficients, which are more easily influenced by recidivism base rates. The sexual recidivism rate over an average follow-up period of 4 to 5 years was 13.7%. The violent recidivism rate, including both sexual and nonsexual violence, was 14.3%. The general recidivism rate was 36.2%. Risk factors for sexual recidivism included deviant sexual preference ($d = .30$) and antisocial orientation ($d = .23$). Antisocial orientation was the biggest risk factor for violent nonsexual ($d = .51$), violent ($d = .54$) and general recidivism ($d = .52$).

Hanson and Morton-Bourgon (2005) expanded upon an earlier meta-analysis by the same research group (Hanson & Bussière, 1998). Hanson and Bussière examined 28,972 sex offenders in 61 studies and subdivided recidivism by original offense type. The sexual recidivism rate over an average follow-up period of 4 to 5 years was 13.4% overall, 18.9% for rapists, and 12.7% for child molesters. The nonsexual violent recidivism rate was 12.2% overall, 22.1% for rapists, and 9.9% for child molesters. The
general recidivism rate was 36.3% overall, 46.2% for rapists, and 36.9% for child molesters. Risk factors for recidivism included failure to complete treatment, sexual deviancy, and general criminological factors (e.g., age and total prior offenses).

Cann, Falshaw, and Friendship (2004) measured the reconviction rates of all 419 adult male sex offenders released from prison in England and Wales in 1979, for a follow-up period of 21 years. The reconviction rate for general sexual offenses was 24.6%, for violent offenses 21.7%, and for general offenses 61.8%. Of the offenders who committed a sexual reoffense during the follow-up period, 35.9% did so after a period of five years or longer. The follow-up period of this study was several times longer than many comparable studies, which likely accounts for the substantially higher recidivism estimates.

Working with a different population, Maletzky and Steinhauser (2002) examined treatment failure rates of 7,275 sex offenders who entered cognitive-behavioral treatment. Treatment failure was defined as one of the following four scenarios: (a) self-report of deviant sexual behavior, (b) deviant physiological arousal post-treatment, (c) deception on sexual questions while taking a polygraph, or (d) a new sex crime charge. Offenders were assigned to one of the following groups: child molesters/female victim \((n = 2,196)\), child molesters/male victim \((n = 765)\), heterosexual pedophiles \((n=1,011)\), homosexual pedophiles \((n = 1,251)\), exhibitionists \((n = 1,604)\), and rapists \((n = 448)\). Maletzky and Steinhauser differentiated between more opportunistic child molesters and pedophiles, assigning to the latter group offenders who demonstrated either an arousal preference for children or a predatory style of offending. The former group consisted of those offenders who did not fulfill either of these criteria and who tended to be more opportunistic in
their sexual offending. After 5 years of follow-up, the failure rate was 4.2% for child molesters/female victim, 7.6% for child molesters/male victim, 9.3% for heterosexual pedophiles, 15.8% for homosexual pedophiles, 12.8% for exhibitionists, and 15.6% for rapists. The overall failure rate was 12.2%, and offenders in some groups continued to be at significant risk for recidivism beyond the 5-year follow-up period. Dropping out of treatment was associated with increased risk of reoffense: 1.5% of treatment completers were eventually charged with a new sex crime, versus 8.1% of treatment dropouts.

These studies are examples of the thorough research that has been conducted into recidivism among sex offenders. Large sample sizes, long follow-up periods, and thorough documentation by legal and criminal justice systems lend credibility to the conclusions drawn by researchers. However, the results of several studies may be interpreted as cautions to unquestioning acceptance of the traditional research design of recidivism studies. For example, Corbett, Patel, Erikson, and Friendship (2003) tested the hypothesis that, for some sex offenders, reconvictions on charges of nonsexual violence sometimes involved criminal acts that in reality were sexually motivated. The authors coded details of reoffenses documented in the case files of 104 adult male sex offenders who had been reconvicted of a nonsexually violent charge and concluded that 12% of the violent reoffenses were in fact sexually motivated. Similarly, Falshaw, Bates, Patel, Corbett, and Friendship (2003) examined the effect of expanding the traditional sexual reconviction outcome measure to include other offense-related sexual behaviors. The authors defined such behaviors as anything indicative of a previous sexual abuse pattern, be it legal or illegal (e.g., a convicted child molester spending significant amounts of time
outside a school). In the sample of 173 sex offenders, broadening the outcome definition in this manner increased the sexual reconviction rate 5.3 times.

In present study, I focused on predicting future sexual recidivism. The studies discussed previously suggest that a substantial minority of sex offenders go on to commit similar crimes in the future. Legal and criminal justice officials attempt to identify future reoffenders in order to provide them with treatment, keep them contained, and restrict their access to potential victims. Such identification requires knowledge of the risk factors of recidivism. I will discuss research into those risk factors in the following sections.

*Role of Antisocial and Psychopathic Traits in Recidivism*

Antisocial Personality Disorder (APD) is a mental disorder described in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision* (*DSM-IV-TR*; American Psychiatric Association, 2000). This disorder is characterized by disregard for laws, rules, and the rights of others, beginning in adolescence and continuing into adulthood. Although the diagnostic criteria for APD include emotional components, such as lack of remorse for wrongdoing, APD is primarily defined in behavioral terms. The majority of individuals who have been convicted of violent crimes fulfill the criteria for this diagnosis.

APD is frequently confused with psychopathy, a condition that is not described in the *DSM-IV-TR* but that can be measured using several assessment instruments (Hare, 2003). Psychopathy is a condition involving a lack of guilt and empathy and a tendency to behave callously and grandiosely. It differs from APD in that its criteria include inner, emotional experience rather than simply overt behavior. The presence of psychopathy
does not depend upon a violent or criminal history, and conversely the majority of people who have been convicted of violent crimes cannot be accurately classified as psychopaths.

An extensive body of research supports the conclusion that both antisocial and psychopathic traits are significant risk factors for recidivism among sex offenders. Many of these studies included the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 2003) as a measure of psychopathic traits. The PCL-R is a rating scale with a range of 0-44. Scores of 30 to 44 indicate the presence of psychopathy, and scores of 21 to 29 indicate elevated psychopathic traits. The PCL-R has been the subject of a large number of empirical studies and several meta-analyses and reviews (e.g., Hemphill, Hare, & Wong, 1998; Salekin, Rogers, & Sewell, 1998). The instrument’s power in predicting recidivism and violence among forensic samples is well established. For example, Hemphill et al. report that over a 1-year follow-up period, criminals who received high scores on the PCL-R were approximately four times as likely to violently recidivate as nonpsychopaths. In some cases the PCL-R has out-performed actuarial risk assessment instruments, making it a powerful tool for predicting violence and recidivism (Barbaree, Seto, Langton, & Peacock, 2001).

Porter et al. (2000) examined the prevalence of high PCL-R scores in a sample of 329 offenders, 228 of whom had committed at least one sex offense. Groups that included relatively large percentages of psychopaths included adult rapists (35.9%), offenders who had both raped adults and molested children (64.0%), and nonsexual offenders (34.0%). Groups that included relatively small percentages of psychopaths included intrafamilial child molesters (10.8%), extrafamilial child molesters (6.3%), and mixed intrafamilial
and extrafamilial child molesters (6.3%). Additionally, Hanson and Morton-Bourgon (2005) found in their meta-analysis of sex offender recidivism in 82 studies that PCL-R scores were a significant risk factor for recidivism.

Serin, Mailloux, and Malcolm (2001) examined the role of psychopathy and deviant sexual arousal in general recidivism in a sample of 68 child molesters and rapists. Psychopathy was assessed using the PCL-R, and deviant arousal was assessed using phallometry. After a 7-year follow-up period, survival analysis indicated that offenders with high PCL-R scores reoffended more quickly ($M = 48$ months) than did offenders with low PCL-R scores ($M = 69$ months). High and low deviant arousal did not differentially predict time to reoffense ($M = 59$ and 61 months, respectively). When the researchers combined high/low PCL-R scores and high/low deviant arousal into four permutations, only high PCL-R/high arousal ($M = 37$ months) and low PCL-R/high arousal ($M = 79$ months) differed significantly from each other. These results suggest that PCL-R scores, but not deviant arousal, are significantly predictive of time to reoffense, though it should be noted that these results were drawn from a relatively small sample size.

Rice and Harris (1997) examined the role of psychopathy and deviant sexual arousal in violent and sexual recidivism in a sample of 288 sex offenders. Survival analysis indicated that the combination of high PCL-R scores and deviant sexual arousal significantly predicted sexual recidivism but not violent recidivism. More specifically, the violent recidivism failure rate for psychopaths was significantly greater than of nonpsychopaths. The sexual recidivism rate for psychopaths with deviant arousal was
significant greater than that of psychopaths without deviant arousal and nonpsychopaths with or without deviant arousal.

Firestone, Bradford, Greenberg, and Serran (2000) examined the relationship between PCL-R scores and deviant sexual arousal in a sample of 156 incest offenders, 260 extrafamilial child molesters, and 123 rapists. Psychopathy tended to be highest among rapists (average PCL-R score of 25.17) than among incest offenders ($M = 18.64$) and extrafamilial child molesters ($M = 18.32$). There was a significant correlation between psychopathy and three out of four scales of deviant arousal in extrafamilial child molesters ($r = .16$ to $.30$). However, there was no significant correlation between psychopathy and deviant arousal in incest offenders and rapists.

In two studies, researchers examined both the PCL-R and the Antisocial Features (ANT) scale of the Personality Assessment Inventory (PAI; Morey, 1991) as measures of antisocial and psychopathic traits among inmates convicted of sexual offenses. The results suggested that high scores on both measures were moderately associated with misbehavior in prison (Buffington-Vollum, Edens, Johnson, & Johnson, 2002; Edens, Buffington-Vollum, Colwell, Johnson, & Johnson, 2002). Rosenberg, Abell, and Mackie (2005) found that high PCL-R scores among adult male child molesters were associated with the use of physical violence against victims.

In sum, antisocial and psychopathic traits have been found to be significantly predictive of future criminal and violent behavior when considered in isolation. These characteristics have also been found to be moderately related to future sexual reoffense among incarcerated sex offenders. One may therefore conclude that assessment of antisocial and psychopathic traits is an important component of sex offender recidivism.
and hypothesize that existing sex offender risk assessment tools that do not incorporate assessment of such traits may be improved by the addition of such data.

**Actuarial Prediction of Sex Offender Recidivism**

Given the prevalence, impact, and recidivism rates associated with sex offenses, it is essential that professionals who work with sex offenders have effective tools for predicting future behavior. The use of empirically derived, actuarial risk assessment instruments as alternatives to clinical judgment of risk has increased markedly in recent years (Janus & Prentky, 2003). As Janus and Prentky noted, such instruments were constructed based upon the strength of statistical correlations between risk factors and outcomes, rather than upon theories of human behavior or the professional judgment of experts. Conclusions drawn using such actuarial instruments may be considered relatively objective. Because of this, such conclusions often form a key part of legal and criminal justice decisions related to sex offenders, including those in, criminal trials, correctional management, conditional release, civil commitment, and psychological treatment.

There is substantial empirical evidence that, in general, actuarial prediction instruments perform as well as or better than unstructured clinical judgment (Grove & Meehl, 1996). Support for sex offense-specific actuarial instruments is more moderate than for other types of actuarial instruments (Becker & Murphy, 1998), but this has not deterred forensic psychologists and other forensic professionals from widely adopting them. Proponents of the use of actuarial risk assessment instruments with sex offenders have asserted that these tools are useful and ethical as long as their conclusions are phrased concretely and appropriate disclaimers are provided (Janus & Meehl, 1997; Janus & Prentky, 2003). Seto (2005) found sufficient commonality among popular actuarial
instruments such that predictive accuracy was not improved by combining results of several such instruments when making risk predictions. Some research even suggests that certain actuarial risk assessment instruments are resistant to the effects of base rate changes in offenses (Doren, 2004), though this assertion has been challenged (Mossman, 2006; Wollert, 2006).

There has also been substantial criticism of actuarial risk assessment instruments, based on a variety of arguments. For example, several authors have noted that the moderate accuracy of sex offense-specific instruments may be insufficient in many legal situations (Craig, Browne, & Stringer, 2003; Rogers & Jackson, 2005). Sjöstedt and Grann (2002) concluded that some actuarial tools may be better suited to predicting common but less severe offenses than rare but more severe offenses. DeClue (2005) warned of the danger of the fundamental attribution error when making clinical decisions about risk, and Glancy (2006) noted that ostensibly objective instruments may be scored differently by different evaluators. Hart, Michie, and Cooke (2007) concluded that some common actuarial instruments have confidence intervals for individual-level prediction that are so wide as to negate the instruments’ utility.

The most recent controversy concerning actuarial risk assessment instruments is their general reliance upon static (i.e., historical) risk factors as opposed to dynamic or changeable risk factors (e.g., response to treatment). Craig, Browne, Stringer, and Beech (2005) reviewed literature related to static and dynamic risk factors and concluded that, although there is increasing agreement within the field about which static factors account for variance in recidivism, there is a paucity of research about dynamic risk factors. Other authors have echoed the call for further research into dynamic risk factors (Craissati &
Beech, 2006; DeClue, 2005; Proulx et al., 1997), particularly those related to psychological treatment (Craig et al., 2003; Hanson, 1998; Studer & Reddon, 1998).

I next discuss one particular actuarial risk assessment instrument, the Static-99. This instrument is one of the most widely-used actuarial risk assessment scales for predicting sexual recidivism among sex offenders (Hanson & Thornton, 1999). In the current study, I tested the hypothesis that the Static-99’s predictive ability would be improved with the addition of data from the Personality Assessment Inventory (PAI; Morey, 1991).

**Static-99**

The Static-99 is a widely-used, 10-item, actuarial risk assessment instrument designed for use with adult male sex offenders (Hanson & Thornton, 1999, 2000; see Table 1). A clinician or appropriately trained nonclinician rates the items and sums them to place the evaluatee in a sexual offense recidivism risk category. The coding rules for the Static-99 were revised in 2003 (Harris, Phenix, Hanson, & Thornton, 2003). The instrument’s authors have released an updated version, the Static-2002 (Hanson & Thornton, 2003), but it remains questionable whether this new instrument will be adopted as a replacement for the Static-99 given the earlier instrument’s popularity, ease of use, and relatively large body of empirical research supporting it (Helmus & Hanson, 2007).
<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Codes</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Young at time of evaluation</td>
<td>Aged 25 or older</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Aged 18-24.99</td>
<td>1</td>
</tr>
<tr>
<td>2. Ever lived with a lover for at least two years?</td>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>3. Index nonsexual violence – any convictions</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>4. Prior nonsexual violence – any convictions</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>5. Prior sex offenses</td>
<td>Charges</td>
<td>Convictions</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>1-2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3-5</td>
<td>2-3</td>
</tr>
<tr>
<td></td>
<td>6+</td>
<td>4+</td>
</tr>
<tr>
<td>6. Prior sentencing dates (excluding index)</td>
<td>3 or less</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4 or more</td>
<td>1</td>
</tr>
<tr>
<td>7. Any convictions for noncontact sex offenses</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>8. Any unrelated victims</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>9. Any stranger victims</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>10. Any male victims</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score</th>
<th>Risk Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 1</td>
<td>Low</td>
</tr>
<tr>
<td>2, 3</td>
<td>Moderate-Low</td>
</tr>
<tr>
<td>4, 5</td>
<td>Moderate-High</td>
</tr>
<tr>
<td>6+</td>
<td>High</td>
</tr>
</tbody>
</table>
Hanson and Thornton (2000) normed the Static-99 using a sample of 1,301 adult male sex offenders from Canada and the United Kingdom. This overall sample was drawn from four smaller samples, each with a different average follow-up period. The first group (minimum \( n = 344 \)) was followed for an average of 4 years, the second (minimum \( n = 191 \)) for an average of 23 years, the third (minimum \( n = 142 \)) for an average of 10 years, and the fourth (minimum \( n = 531 \)) for an average of 16 years. The first and fourth groups were drawn from correctional facilities, and the second and third groups were drawn from secure psychiatric hospitals. The individuals had been convicted of a variety of contact and noncontact sex offenses.

Using receiver operating characteristic (ROC) analysis, Hanson and Thornton calculated the Static-99’s area under the curve (AUC) as .70 for sexual reoffending and .69 for violent reoffending. This result suggests that the Static-99 is a moderately powerful predictive instrument. The Static-99 manual (Harris et al., 2003) lists interrater reliability scores ranging from .80-.96. Subsequent researchers have reported sexual recidivism AUCs for the Static-99 as high as .78 (Allan, Dawson, & Allan, 2006) and as low as .52 (Craig, Browne, Beech, & Stringer, 2004), violent recidivism AUCs as high as .74 (Sjöstedt & Långström, 2001) and as low as .60 (Craig et al., 2004), and general recidivism AUCs as high as .65 (Harris et al., 2003) and as low as .58 (Craig et al., 2004). Generally, these and other replications have been consistent with the authors' original

---

1 Receiver operating characteristic (ROC) analysis is a statistical technique for calculating an instrument’s sensitivity when making binary classifications or predictions, such as true/false or reoffend/nonreoffend. Graphing the results of an ROC analysis yields a curve, and the area under the curve (AUC) is a measure of instrument sensitivity. AUCs of a moderate level (.65 to .80) are frequently described as acceptable in forensic psychology studies (e.g., Allan, Dawson, & Allan, 2006; Sjöstedt & Långström, 2001; Harris et al., 2003). Scores above .80 are unusual.
data (Langton et al., 2007; Looman, 2006; Sjöstedt & Grann, 2002; Sjöstedt & Långström, 2001; Stadtland et al., 2005). These results support the conclusion that the Static-99 is a moderately powerful predictive instrument. The predictive performance of the Static-99 shows some evidence of resistance to interference from changes in crime base rates (Doren, 2004).

Barbaree, Langton, and Peacock (2006) examined the factor structure of items from several widely-used risk assessment instruments. Their analysis yielded six factors, including antisocial behavior, sexual abuse of a child by the offender, persistence, detached predatory behavior, offender status as young and single, and male victim(s). The Static-99 was significantly correlated with the scores on five of the six factors found in popular instruments, including antisocial behavior ($r = .35$), persistence ($r = .59$), detached predatory behavior ($r = .52$), young and single ($r = .23$), and male victim(s; $r = .25$). The instrument was not significantly correlated with child sexual abuse ($r = -.01$).

Most researchers comparing the Static-99 to other, established risk assessment instruments have suggested that the Static-99 performs as well as or better than those other instruments (Barbaree et al., 2001; Bartosh, Garvy, Lewis, & Gray, 2003; Craissati & Beech, 2005; Hanson & Thornton, 2000; Nunes, Firestone, Bradford, Greenberg, & Broom, 2002; Sjöstedt & Grann, 2002; Sjöstedt & Långström, 2001; Stadtland et al., 2005). For example, Hanson and Thornton's study using the Static-99's norming sample ($n = 1,301$) revealed that the Static-99's predictive accuracy (AUC = .70 for sexual recidivism and AUC = .69 for violent recidivism) slightly exceeded the predictive accuracy of either of its component scales, the Rapid Risk Assessment of Sexual Offender Recidivism (RRASOR; Hanson, 1997; sexual AUC = .68, violent AUC = .65).
and the Structured Anchored Clinical Judgment (SACJ-Min; Grubin, 1998; sexual AUC = .69, violent AUC = .67). Barbaree et al. compared the predictive accuracy of seven risk assessment tools for 215 sex offenders: the Static-99, Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993), Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998), RRASOR, Minnesota Sex Offender Screening Tool-Revised (MnSOST-R; Epperson, Kaul, & Hesselton, 1998), Multifactorial Assessment of Sex Offender Risk for Recidivism (MASSOR; a structured interview that was first evaluated in this study), and PCL-R. The researchers found that the STATIC-99, RRASOR, VRAG, and SORAG outperformed the other instruments in predicting sexual, serious, and general recidivism and, although some of those four instruments performed somewhat better at certain types of recidivism, overall they performed similarly (i.e., generally in the range of AUC = .65 to .77).

Other researchers found marginally better performance with other instruments for certain outcomes but still reported generally positive results for the Static-99 (Allan et al., 2003; Harris et al., 2003; Looman, 2006). Gentry, Dulmus, and Theriot (2005) compared the Static-99 with the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995) and found that 63.3% of cases were classified as higher risk when the Static-99 was used. This difference in classification suggests that the Static-99 is a more conservative risk assessment instrument than the LSI-R. It should be noted that Gentry et al. did not assess recidivism.

Several researchers have tested modifications to the Static-99's norms, structure, and interpretation in order to increase its predictive power. Nunes et al. (2002) were able to slightly increase the instrument's predictive power by adding phallometric data.
Sjöstedt & Grann (2002) found that using more complex, nondichotomous outcome variables (taking into account the imminence, frequency, nature, and severity of the reoffense, rather than simply the fact that a reoffense occurred) had mixed effects on the Static-99’s AUC. Craig, Thornton, Beech, and Browne (2007) found that the instrument's items could be meaningfully subdivided into three subscales: sexual deviance, general criminality, and immaturity. This result contrasts with the five meaningful factors described in Barbaree et al. (2006) but may reflect a difference in methodology: Barbaree et al. performed their factor analysis on items from five risk assessment instruments, whereas Thornton et al. performed their factor analysis only on Static-99 items. Seto (2005) found no benefit to combining various popular actuarial risk assessment instruments with the Static-99, and Craig et al. (2004) found little benefit to adding personality data from the Special Hospitals Assessment of Personality and Socialization (SHAPS) to the Static-99. Craissati and Beech (2005) proposed the addition of data about childhood events, and Allan et al. (2006) proposed developing separate scales for violent and nonviolent offenders. In summary, although several researchers have tested modifications to the Static-99, improvements in predictive ability have been marginal at best, studies have not been replicated, and modifications to the Static-99 and have not been widely adopted.

Several researchers have examined the utility of the Static-99 among sex offenders from countries other than Canada, the United States, and the United Kingdom, and have reported mixed results. Several of these studies have suggested that the results of the instrument generalize well to general (i.e., primarily Caucasian) European sex offender populations (Ducro & Pham, 2006; de Vogel, de Ruiter, van Beek, & Mead,
There is also evidence that the instrument can be meaningfully used with Japanese sex offenders (Sudo, Sato, Obaba, & Yamagami, 2006). However, a study by Långström (2004) suggested that the instrument’s results may not generalize well to non-Caucasian Europeans, and Allan et al.’s (2006) study concluded with the advice that the Static-99 not be used with indigenous Australians without further research.

Several researchers have raised additional criticisms of the Static-99. These criticisms include unacceptably broad confidence intervals for individual-level prediction (Hart et al., 2007), lack of dynamic risk factors (Craig et al., 2003, 2005), sensitivity to changes in crime base rates (Mossman, 2006), and limited utility for high-stakes sexually violent predator litigation (Rogers & Jackson, 2005). Soothill, Harman, Francis, and Kirby (2005) suggested that the Static-99 may classify offenders at higher risk more accurately than offenders at lower risk, whereas Sjöstedt and Grann (2002) suggested the opposite. Glancy (2006) warned that the Static-99's predictive accuracy may be decreased by inconsistent scoring.

In summary, the Static-99 is widely used and performs well relative to alternative actuarial risk assessment instruments. However, like many other actuarial risk assessment instruments, its predictive ability is in the moderately accurate range, and given that high-stakes decisions are routinely justified with the results of the Static-99, more research should be done to improve its predictive accuracy (Janus & Prentky, 2003). In the next section, I describe two scales of a personality inventory that address antisocial and psychopathic traits, which previous research has suggested are predictive of future
violence and criminality. In the present study, I tested the hypothesis that adding these data to Static-99 data would result in more accurate predictions of sexual recidivism.

Personality Assessment Inventory

The Personality Assessment Inventory (PAI; Morey, 1991) is a 344-item, self-report empirically derived measure of personality and psychopathology. It was developed using DSM-III-R diagnostic criteria. It yields 4 validity scales, 11 clinical scales, 5 treatment scales, and 2 interpersonal scales. The PAI was normed on a sample of adults and may be completed by test-takers with a fourth-grade reading level. Because it represents a psychometric improvement over comparable instruments, the PAI has been widely adopted in forensic settings, where challenges to instrument reliability, validity, and generalizability often arise during the process of litigation (Edens, Cruise, & Buffington-Vollum, 2001; Morey & Quigley, 2002). In this study, I focused on the use of two PAI scales, Antisocial Features (ANT) and Aggression (AGG), which may be particularly useful in predicting risk of recidivism among sex offenders.

Antisocial Features (ANT)

Morey (1991) developed the ANT scale to measure "personality and behavioral features relevant to the constructs of antisocial personality and psychopathy" (p. 18). This clinical scale is made up of three subscales, including Antisocial Behaviors (ANT-A), Stimulus-Seeking (ANT-S), and Egocentricity (ANT-E). Research using the PCL-R, which has been discussed previously, has suggested that the latter two interpersonal and affective domains are essential domains of the construct of psychopathy, and that the former behavioral domain may be a common byproduct of such traits (Hart, 2008). ANT-A is theoretically related to the DSM-IV-TR diagnostic criteria for Antisocial Personality
Disorder, which draw heavily upon an individual’s history of criminal behavior and is significantly correlated with the antisocial lifestyle factor of the PCL-R (Edens, Hart, Johnson, Johnson, & Olver, 2000). ANT-S is theoretically related to the antisocial behavior factor of psychopathy as defined by Hare (2003), and Edens et al. found that it correlated significantly with this second factor of the PCL-R. ANT-E is theoretically related to the affective/interpersonal dimension of psychopathy as defined by Hare, though Edens et al. found that it did not correlate significantly with this first PCL-R factor.

Several researchers have examined the extent to which ANT and the PCL-R measure comparable constructs. Buffington-Vollum et al. (2002) and Edens et al. (2002) concluded that the instruments have substantial overlap but also measure a significant amount of unique information. Edens et al. (2000) concluded that ANT more closely captured the behavioral aspects of psychopathy than the condition’s interpersonal and affective symptoms. They reported a correlation of .40 between ANT and the PCL-R total score. ANT-A and ANT-S loaded significantly on the PCL-R total score and Factor 2 (antisocial lifestyle), whereas ANT-E did not load significantly on the PCL-R total score or on either factor (emotional/interpersonal or antisocial lifestyle). Douglas, Guy, Edens, Boer, and Hamilton (2007) examined models of PCL-R psychopathy derived from PAI profiles. The results suggested that models based on ANT were better at screening out psychopathy than they were at identifying high levels of psychopathy. The researchers cautioned against using the PCL-R and ANT interchangeably. The results of these studies suggest that some of the ability of the PCL-R to predict recidivism may translate to ANT but that research should be done to test this hypothesis before ANT is
used for this purpose in clinical contexts. The results also suggest that ANT may have unique predictive ability over and above PCL-R data.

Research suggests that ANT is significantly predictive of institutional violence and rule-breaking in forensic populations. Buffington-Vollum et al. (2002), Caperton, Edens, and Johnson (2004), and Edens et al. (2002) all found that ANT predicted both aggressive and nonaggressive infractions by incarcerated sex offenders, though there was some evidence that ANT’s predictive ability decreased when predicting physically violent infractions. In contrast to these results, Walters (2007) found that ANT was a poor predictor of similar institutional adjustment in a sample of general inmates. In their 2001 review of literature related to the PAI’s use in forensic settings, Edens et al. concluded that ANT significantly predicted both institutional misconduct and criminal recidivism.

**Aggression (AGG)**

AGG is a treatment scale, not a clinical scale like ANT (Morey, 1991). As such, AGG was designed for risk assessment and treatment planning, rather than as a diagnostic aid. Individuals who score highly on this scale tend to have trouble controlling their anger and to be perceived by others as hostile. Morey found that in the norming sample, high scores were correlated with both verbally and physically aggressive behavior. AGG is composed of three subscales, including Aggressive Attitude (AGG-A), Verbal Aggression (AGG-V), and Physical Aggression (AGG-P). The PAI manual provides two different correlations between ANT and AGG. In the clinical norming sample ($n = 1,246$), the correlation was .61. In the census-matched normative sample ($n = 1,000$), the correlation was .55.
There is evidence that AGG is a useful predictive scale for forensic populations, though less research has been conducted using this scale than has been conducted using ANT. Caperton et al. (2004) found that AGG was somewhat predictive of institutional misbehavior, though it did not offer incremental validity beyond ANT. Walters (2007) found that AGG was significantly predictive of both aggressive and nonaggressive institutional misconduct among general inmates. Edens et al. (2001) concluded that AGG significantly predicted both institutional misconduct and criminal recidivism.

Rationale for Current Study

In this study, I examined the hypothesis that the PAI ANT and AGG scales would add incremental validity to the Static-99 in predicting sexual recidivism among adult male sex offenders. Current research suggests that the Static-99, ANT, and AGG all have some utility individually as risk assessment tools and also suggests that they may account for unique variance in such predictions of risk (Edens et al., 2001; Hanson & Thornton, 1999). The Static-99 does not directly address aspects of antisocial and psychopathic traits that may predict sexual recidivism, and the ANT and AGG scales do not directly assess sex offense-specific aspects of sexual recidivism. Furthermore, there is evidence that the well-researched predictive utility of the PCL-R may be replicated, at least in part, by ANT and that ANT may additionally contribute unique variance in measuring antisocial personality. Given the prevalence of Static-99 and PAI usage in forensic settings, any incremental improvement may be useful to those who work with sex offenders.

In the MacArthur Violence Risk Assessment study, Monahan et al. (2001) proposed a model for integrating risk assessment data from different instruments, each of
which measures distinct but related information relevant to violent behavior. They found that using Iterative Classification Tree statistical methodology to combine the unique predictive aspects of several instruments resulted in improved predictive accuracy over any of the constituent instruments. That study set a precedent for testing one method of statistically combining risk assessment data from different sources. In the current study, I test an alternative approach described below. Several researchers have examined similar strategies using the Static-99, with mixed results (Craig et al., 2004; Nunes et al., 2002; Seto, 2005). It is theoretically plausible that the Static-99, ANT, and AGG measure distinct but related information relevant to the commission of future sex offenses. I was unable to locate previous research examining the integration of Static-99 and PAI data in this manner, and therefore I performed an analysis to test the hypothesis that such a combination would result in improved predictive accuracy of sexual rearrest and reconviction.
METHOD

Sample Characteristics

Subjects in this study were 72 adult males receiving evaluation or treatment for sexual deviancy at the Center for Behavioral Intervention, a community-based sex offender treatment agency in Beaverton, Oregon. Subjects included individuals who had been convicted of sex offenses, individuals who had been charged with sex offenses but who were not yet adjudicated, individuals under investigation by child protective services who had previously been legally investigated for sex offenses, and individuals who were not under investigation or oversight at the time of evaluation but had previously been legally investigated for sex offenses (see Table 2 for percentages). The subjects presented with a variety of sexual offense types (e.g., rape, child molestation, non-contact sex offenses). The offenders in this sample varied across a number of sociodemographic factors (see Table 2 for descriptive statistics). This sample is fairly unique in that it focused on individuals in community-based treatment, rather than solely upon incarcerated sex offenders as is typical in most studies. Thus, in addition to testing the incremental validity hypotheses, this study provides data on the generalizability of the Static-99 and the PAI’s ANT and AGG scales in populations different from those of the scales’ standardization samples.
Table 2
Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>39.13 (12.41)</td>
</tr>
<tr>
<td>% (n)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>6.9% (5)</td>
</tr>
<tr>
<td>Asian</td>
<td>1.4% (1)</td>
</tr>
<tr>
<td>Bi- or multiracial</td>
<td>2.8% (2)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>77.8% (56)</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>11.1% (8)</td>
</tr>
<tr>
<td>Crime</td>
<td></td>
</tr>
<tr>
<td>Abuse of child and adult</td>
<td>2.8% (2)</td>
</tr>
<tr>
<td>Child sexual abuse</td>
<td>66.7% (48)</td>
</tr>
<tr>
<td>Exposure</td>
<td>9.7% (7)</td>
</tr>
<tr>
<td>Rape</td>
<td>16.7% (12)</td>
</tr>
<tr>
<td>Sodomy</td>
<td>4.2% (3)</td>
</tr>
<tr>
<td>Status at Evaluation</td>
<td></td>
</tr>
<tr>
<td>Charges dropped or case dismissed</td>
<td>2.8% (2)</td>
</tr>
<tr>
<td>Convicted and awaiting sentence</td>
<td>9.7% (7)</td>
</tr>
<tr>
<td>Convicted and sentenced</td>
<td>34.7% (25)</td>
</tr>
<tr>
<td>Correctional sanction</td>
<td>1.4% (1)</td>
</tr>
<tr>
<td>Child protective investigation</td>
<td>9.7% (7)</td>
</tr>
<tr>
<td>Pre-adjudication</td>
<td>41.7% (30)</td>
</tr>
<tr>
<td>Sexual arrest during follow-up</td>
<td>8.3% (6)</td>
</tr>
<tr>
<td>Sexual conviction during follow-up</td>
<td>2.8% (2)</td>
</tr>
<tr>
<td>Static-99</td>
<td>2.44 (1.86)</td>
</tr>
<tr>
<td>PAI-ANT</td>
<td>52.76 (9.63)</td>
</tr>
<tr>
<td>PAI-AGG</td>
<td>46.15 (10.39)</td>
</tr>
</tbody>
</table>

1 For these cases, there were sufficient prior offenses to code the Static-99.
Instruments

**Static-99**

As discussed above, the Static-99 is a widely-used, 10-item, actuarial risk assessment instrument designed for use with adult male sex offenders (Hanson & Thornton, 1999, 2000; see Table 1). A clinician or appropriately trained nonclinician rates the items and sums them to place the evaluatee in a recidivism risk category. The range of possible scores is 0 to 12, and the recidivism risk categories include low (0, 1), moderate-low (2, 3), moderate-high (4, 5), and high (6+). Hanson and Thornton normed the Static-99 using a sample of 1,301 adult male sex offenders from Canada and the United Kingdom. They calculated the Static-99’s AUC as .70 for sexual reoffending and .69 for violent reoffending. This result suggests that the Static-99 is a moderately powerful predictive instrument. The Static-99 manual (Harris et al., 2003) lists interrater reliability alphas ranging from .80-.96.

**Personality Assessment Inventory**

As discussed earlier, the PAI (Morey, 1991) is a 344-item, self-report measure of personality and psychopathology. The scale was developed in a primarily empirical manner based on modern diagnostic criteria and personality theory. The PAI yields 4 validity scales, 11 clinical scales, 5 treatment scales, and 2 interpersonal scales. In this study, I focused on the use of two PAI scales: Antisocial Features (ANT, a clinical scale) and Aggression (AGG, a treatment scale), which may be particularly useful in predicting risk of recidivism among sex offenders. ANT and AGG scores of 59 or below are considered low, scores of 60 to 69 are considered medium, and scores of 70+ are considered high. I decided a priori that subjects' PAI profiles would be excluded as
invalid if two or more of the validity scales were in the high range as defined by the manual (Inconsistency = 73+, Infrequency = 75+, Negative Impression Management = 92+, Positive Impression Management = 68+; Morey, 1991). According to that exclusion criterion, no profiles were excluded as invalid. Seven subjects had one validity scale in the high range.

The PAI has demonstrated strong psychometric reliability and validity in a variety of populations, including individuals in forensic settings (Edens et al., 2001). The instrument’s manual (Morey, 1991) lists internal consistency alphas of .81, .86, and .82 for norming samples of normative, clinical, and collegiate individuals. The instrument’s 22 scales correlate highly with established personality and psychopathology assessments (e.g., MMPI, NEO-PI), while at the same time demonstrating greater adherence to recent changes in diagnostic criteria. Norming data from forensic samples, including incarcerated individuals and people who received high psychopathy scores, indicate that the PAI is a valid and useful instrument for use with forensic populations. This conclusion has been supported by subsequent research (Buffington-Vollum et al., 2002; Caperton et al., 2004; Douglas et al., 2007; Edens et al., 2000, 2001, 2002; Morey & Quigley, 2002).

Procedure

Permission from the Institutional Review Boards of Pacific University and the Oregon Department of Corrections was obtained prior to data collection, which was conducted through a review of archival data. The subjects had already completed the PAI as part of an intake assessment battery. The Static-99 was scored by graduate-level research assistants based on self-reported offense histories, psychological evaluation
reports, and legal records in the case archives. Sexual recidivism data were obtained from the Oregon Department of Corrections. For subjects who had not been convicted of the offense that prompted the psychological evaluation, later conviction for that offense was not counted as recidivism. The average follow-up period following the evaluation was 5.2 years and the follow-up times ranged from 1 to 10 years.
RESULTS

A series of correlation analyses were conducted to determine the degree to which Static-99 scores, ANT scores, and AGG scores were correlated in this sample. Next, two one-way analyses of variance (ANOVA) were conducted to determine whether subjects with different ANT and AGG groupings had significant different Static-99 scores. Two binary logistic regression analyses were conducted to determine whether Static-99 scores, ANT scores, and AGG scores accounted for unique variance in sexual rearrest and sexual reconviction. Finally, eight ROC analyses were performed to determine the predictive validity of the measurement combinations with regard to sexual rearrest and reconviction: Static-99, Static-99 + ANT, Static-99 + AGG, and Static-99 + ANT + AGG on those two outcome variables. Each of these analyses will be discussed in turn.

To construct the combination of Static-99 and PAI scale scores, ANT and AGG were coded as 0 for low scores (59T or below), 1 for medium scores (60T-69T), or 2 for high scores (70T or above). The values of 0-2 were then added to the Static-99 total score. Therefore, Static-99 + ANT ($M = 2.72$, $SD = 2.14$) and Static-99 + AGG ($M = 2.60$, $SD = 2.05$) had a potential range of 0-14 and Static-99 + ANT + AGG ($M = 2.88$, $SD = 2.38$) had a potential range of 0-16. Table 3 displays the distribution of subjects across the three levels of these scales.
Table 3
*Numbers of Subjects in Various Risk Categories across ANT and AGG Levels*

<table>
<thead>
<tr>
<th></th>
<th>ANT</th>
<th>AGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (0T-59T)</td>
<td>57</td>
<td>64</td>
</tr>
<tr>
<td>Medium (60T-69T)</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>High (70T+)</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

The correlations among the Static-99, ANT, and AGG are presented in Table 4. These results suggest that the Static-99 and ANT were moderately correlated, the Static-99 and AGG were weakly correlated, and ANT and ATT were strongly correlated. The weak-to-modest correlation between the Static-99 and the two PAI scales suggested that they measured substantial amounts of unique information.

Table 4
*Correlations between Independent Variables*

<table>
<thead>
<tr>
<th></th>
<th>Static-99</th>
<th>ANT</th>
<th>AGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Static-99</td>
<td>-</td>
<td>.49**</td>
<td>.26*</td>
</tr>
<tr>
<td>2. ANT</td>
<td>-</td>
<td></td>
<td>.67**</td>
</tr>
<tr>
<td>3. AGG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .01, ** p < .05

Two one-way ANOVAs were performed to determine whether subjects with different ANT and AGG groupings had significantly different Static-99 scores. Those Static-99 means and standard deviations are presented in Table 5. The ANOVA for ANT was significant, F (2, 69) = 5.69, MSE = 3.06, p < .01, suggesting that there were significant differences in Static-99 scores across different levels of ANT. Subject status as low, medium, or high on ANT accounted for 14% of the variance in Static-99 score.
Post-hoc Tukey HSD tests of pairwise comparisons of means revealed a significant difference in Static-99 scores between those with low ANT scores and those with medium ANT scores. All other pairwise comparisons were nonsignificant, including the comparison between low and high ANT scores. The small number of subjects with high ANT scores ($n = 5$) may account for the nonsignificance of the comparison of low and high scores, in spite of the fact that the high-scoring group had the same mean ANT score as the medium-scoring group.

The ANOVA for AGG was also significant, $F (2, 69) = 4.32$, $MSE = 3.17$, $p < .05$, suggesting that there were significant differences in Static-99 score between different levels of AGG. Subject status as low, medium, or high on AGG accounted for 11% of the variance in Static-99 score. Post-hoc Tukey HSD tests of pairwise comparisons of means revealed a significant difference in Static-99 scores between those with low AGG scores and those with high AGG scores. All other pairwise comparisons were nonsignificant.

| Table 5 |
| Static-99 Scores across ANT and AGG Groupings |
| ANT | AGG |
| Low (0T-59T) | 2.09 (1.73)$^1$ | 2.28 (1.78)$^1$ |
| Medium (60T-69T) | 3.80 (2.10)$^1$ | 2.80 (1.48) |
| High (70T+) | 3.80 (1.10) | 5.33 (2.31)$^1$ |

$^1$ Pairwise comparison significant at $p < .05$

The results of the binary logistic regressions are presented in Table 6. The first logistic regression tested the hypothesis that Static-99 scores, ANT scores, and AGG
scores accounted for unique variance in sexual rearrest. The second logistic regression tested the hypothesis that Static-99 scores, ANT scores, and AGG scores accounted for unique variance in sexual reconviction. Neither of these analyses yielded significant results, suggesting that in this sample none of the three predictor variables accounted for significant variance in sexual rearrest or sexual reconviction.

Table 6
*Binary Logistic Regressions*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sexual Rearrest B (SE)</th>
<th>Sexual Reconviction B (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99</td>
<td>.08 (.25)</td>
<td>.57 (.39)</td>
</tr>
<tr>
<td>ANT</td>
<td>.69 (.70)</td>
<td>-17.33 (9430.50)</td>
</tr>
<tr>
<td>AGG</td>
<td>1.00 (.27)</td>
<td>-.92 (9927.85)</td>
</tr>
</tbody>
</table>

The results of the ROC analyses are presented in Table 7. These results suggested that neither the Static-99 nor any of the three modifications proposed in the current study (i.e., addition of ANT, AGG, or both) were significantly predictive of sexual rearrest or sexual reconviction during this follow-up period.

Table 7
*Areas under ROC Curves*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Sexual Rearrest</th>
<th>Sexual Reconviction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99</td>
<td>.53</td>
<td>.63</td>
</tr>
<tr>
<td>Static-99 + ANT</td>
<td>.56</td>
<td>.61</td>
</tr>
<tr>
<td>Static-99 + AGG</td>
<td>.55</td>
<td>.62</td>
</tr>
<tr>
<td>Static-99 + ANT + AGG</td>
<td>.57</td>
<td>.60</td>
</tr>
</tbody>
</table>
DISCUSSION

Review of Findings

In the current study, I tested the hypothesis that the addition of data from the ANT and AGG scales of the PAI would improve the predictive power of the Static-99 in a sample of adult male sex offenders evaluated at a community treatment clinic. I did not find support for this hypothesis, because neither the Static-99 nor any of the modifications incorporating ANT or AGG significantly predicted sexual reconviction during the follow-up period. Rearrest for sexual crimes was included as a potentially more sensitive dependent measure of sexual reoffense, but none of the measures showed a predictive relationship with this variable. Also, Static-99 scores were distributed significantly differently across varying levels of ANT and AGG. This differential distribution suggests that the Static-99 is responsive to various levels of risk among community-treated adult male sex offenders. However, the analyses of predictive significance suggest that this community sample may be different enough from the Static-99 norming samples to produce decreased instrument accuracy.

Implications of Findings

The results of this study are largely discrepant with existing literature examining the use of the Static-99 with sex offenders. Although there was a fair amount of variability among studies, the majority of researchers have found that the Static-99 moderately predicted sexual recidivism, which was usually defined as sexual reconviction. This study’s nonsignificant findings may reflect the nature of a more
heterogeneous, community-based sample. A sizeable portion of the subjects did not have a criminal history for sexual offenses compared to most other studies that utilized samples of sex offenders released from prison. Unlike many other studies, the current study did not exclude cases based on psychiatric comorbidity, complexity of offending behavior, or other complicating variables, making the current sample less uniform but more representative of real-world patients. The findings of this study are important in that they suggest that risk assessment measures may exhibit differential utility depending on the nature of the person being evaluated. If the results are replicated in other community-based samples, it suggests some cause for concern when the Static-99 is used in part to make legal decisions about persons accused of committing sexual offenses. If the Static-99 demonstrates little to no utility in assessing risk for sexual re-offense in non-incarcerated samples, then mental health professionals may be doing more harm than good in using the measure in psycholegal evaluations.

Another explanation for the nonsignificant findings is the fact that the base rates of sexual rearrest (8.3%) and sexual reconviction (2.8%) in the sample were low. Although statistical theory suggests that lower base rates of the outcome behavior make prediction more difficult, the reconviction and rearrest percentages are not unlike other samples. Following subjects for a longer follow-up period likely would result in larger base rates of sexual offense reoccurrence. However, other studies with lower base rates of recidivism still led to significant prediction with the Static-99. Given these findings, it is possible that this study indicates reduced utility of the Static-99 with typical community-based populations relative to “cleaner” experimental samples.
Finally, it is possible that a different sample with greater similarity to the Static-99 norming samples would demonstrate one or more of the modifications to be significantly predictive of sexual rearrest or recidivism. Such increased similarity could be achieved by including fewer subjects with serious psychiatric comorbidity, unusual offense details, or sparse or confusing case information. However, it is important to note that the Static-99 and other risk assessment measures are frequently used with these “less clean” types of individuals in psycholegal evaluations (Janus & Prentky, 2003). Including more subjects would also make this study’s sample more similar to the norming sample, although a power analysis conducted a priori suggested that the current sample size of 72 should be sufficient to detect significant statistical trends.

Strengths and Limitations of Current Study

This study’s primary strength is its inclusion of subjects drawn from a community-based sex offender treatment center. This sample, which was filtered through relatively few exclusionary criteria, is likely more representative of typical clinical samples than are more thoroughly screened research samples. This study’s sample provided not only information about my modification to the Static-99, but also about the performance of the unmodified Static-99 in a sample that was less “clean” than the norming sample. In addition, research assistants had access to extensive archival data about each case, allowing them to code the Static-99 based on thorough evaluative information.

Limitations of the current study also resulted primarily from its sample. The sample was community-based and had few exclusionary criteria. Subjects with valid PAIs and sufficient offense histories to code Static-99s were included without
consideration of other factors. This type of sample represents both a strength and a limitation of the current study, in that it provides information about a representative group of community-evaluated sex offenders but also increases error in the analyses. Specifically, such error could potentially have resulted from such factors as the presence or absence of offender mental illness, the variability of demographic and criminal history variables represented, varying sources and types of collateral documents, nonstandardized clinical interviews, or changes in approach or expertise on the part of the clinicians of record over time. It is possible that increasing the sample size would result in more significant findings by increasing the power of the analyses to overcome this increased error.

There are several other possible sources of error in the current study. The Static-99s used in this sample were coded by several researchers, which increased the possibility of scoring error. The researchers used treatment files exclusively to code Static-99s, and any relevant data that were not included in the files were not reflected in the current analyses. Finally, the current sample consisted primarily of Caucasians (77.8%) who were convicted of child sexual abuse (66.7%), limiting the generalizability of the findings to other ethnic and offense type groups.

Directions of Future Study

If the predictive ability of the Static-99 was consistently shown to be decreased in community-based samples, the utility of the instrument to the forensic psychologists would be in question. Future researchers should examine the ability of the Static-99 to predict sexual rearrest and reconviction in different but similarly heterogeneous community-based samples. The inclusion of sexual rearrest as a component of
reoffending is important because such rearrest has a higher base rate than sexual reconviction, which is typically analyzed as the sole proxy of sexual reoffending. A risk assessment instrument that purports to predict sexual reoffense should be able to predict the more sensitive variable of sexual rearrest in addition to the less common outcome of sexual reconviction.

The current study did not indicate that adding data from the ANT and AGG scales of the PAI added predictive ability to the Static-99 alone. It is not clear, however, whether such Static-99 modifications would result in better predictive ability in a sample that more closely resembled the Static-99 norming samples. Given the theoretical plausibility of these modifications, future researchers should examine them empirically. It is also possible that different strategies for combing Static-99 and PAI data would result in more powerful predictions of sexual rearrest and reconviction.
REFERENCES


