Revisiting Risk Assessment with Sexual Offenders: Validation and Extension of the STATIC-99 and Penile Plethysmograph with a Community Sample

Jacqueline Means
Pacific University

Recommended Citation

This Dissertation is brought to you for free and open access by the College of Health Professions at CommonKnowledge. It has been accepted for inclusion in School of Graduate Psychology by an authorized administrator of CommonKnowledge. For more information, please contact CommonKnowledge@pacificu.edu.
Revisiting Risk Assessment with Sexual Offenders: Validation and Extension of the STATIC-99 and Penile Plethysmograph with a Community Sample

Abstract
The ability of the Static-99 and dimensional measurement of penile plethysmograph (PPG) to predict conviction for new general or sexual offenses was assessed with a community sample. Because deviant sexual arousal is theoretically a dynamic risk factor and treatment target, this study proposed that adding the PPG measured deviant arousal to the Static-99 would improve the predictive quality of the actuarial measure. Records for 131 subjects who presented for psychosexual evaluations at a community sex offender treatment facility were examined to collect relevant data and score the Static-99, and conviction information was obtained from the Oregon Department of Corrections. These data were examined using logistic regression to determine the predictive validity of the Static-99, predictive qualities of the PPG, and incremental validity of adding PPG information to the Static-99. Contrary to previous studies examining the Static-99 and PPG separately, results indicated that there were no statistically significant relationship between Static-99 scores, PPG percent deviant arousal, and any new sexual or general convictions. Similarly, the addition of the PPG results to the Static-99 offered no statistically significant increase in predictive abilities. The Static-99 and PPG results performed at chance when predicting recidivism, either generally or sexually, when examined in a community setting. In order to produce more reliable measures of recidivism with offenders released in the community, the Static-99 norms and items, as well as the interpretation and utility of PPG results, must be revisited and refined.

Degree Type
Dissertation

Degree Name
Doctor of Psychology (PsyD)

Committee Chair
Michelle R. Guyton, Psy.D.

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

Third Advisor
Michel Hersen, Ph.D., ABPP

Keywords
Sexual Offending, Risk Assessment, Penile Plethysmograph, Static-99

Subject Categories
Psychiatry and Psychology

This dissertation is available at CommonKnowledge: https://commons.pacificu.edu/spp/130
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
REVISITING RISK ASSESSMENT WITH SEXUAL OFFENDERS:
VALIDATION AND EXTENSION OF THE STATIC-99 AND PENILE
PLETHYSMOGRAPH WITH A COMMUNITY SAMPLE

A DISSERTATION
SUBMITTED TO THE FACULTY
OF
SCHOOL OF PROFESSIONAL PSYCHOLOGY
PACIFIC UNIVERSITY, HILLSBORO, OREGON

BY
JACQUELINE MEANS

IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PSYCHOLOGY

JULY 23, 2010

APPROVED BY THE COMMITTEE:
Michelle R. Guyton, Psy.D.

Genevieve L. Y. Arnaut, Ph.D., Psy.D.

PROFESSOR AND DEAN:
Michel Hersen, Ph.D., ABPP
ABSTRACT

The ability of the Static-99 and dimensional measurement of penile plethysmograph (PPG) to predict conviction for new general or sexual offenses was assessed with a community sample. Because deviant sexual arousal is theoretically a dynamic risk factor and treatment target, this study proposed that adding the PPG-measured deviant arousal to the Static-99 would improve the predictive quality of the actuarial measure. Records for 131 subjects who presented for psychosexual evaluations at a community sex offender treatment facility were examined to collect relevant data and score the Static-99, and conviction information was obtained from the Oregon Department of Corrections. These data were examined using logistic regression to determine the predictive validity of the Static-99, predictive qualities of the PPG, and incremental validity of adding PPG information to the Static-99. Contrary to previous studies examining the Static-99 and PPG separately, results indicated that there were no statistically significant relationship between Static-99 scores, PPG percent deviant arousal, and any new sexual or general convictions. Similarly, the addition of the PPG results to the Static-99 offered no statistically significant increase in predictive abilities. The Static-99 and PPG results performed at chance when predicting recidivism, either generally or sexually, when examined in a community setting. In order to produce more reliable measures of recidivism with offenders released in the community, the Static-99
norms and items, as well as the interpretation and utility of PPG results, must be revisited and refined.

Keywords: Sexual Offending, Risk Assessment, Penile Plethysmograph, Static-99
ACKNOWLEDGEMENTS

I would like to thank a number of people who have encouraged me during the journey through the dissertation process. Firstly, I am grateful to Dr. Michelle Guyton for her guidance and support at every step of this research effort. Erica deserves many accolades for heading the data collection efforts and I am grateful to Cory and Steve Jensen for providing access to 30 years of precious, albeit dusty, data. I would like to thank my colleagues, including Pamela, Christopher, and Kathryn, who contributed greatly to the tedious process of data collection. Thank you to Jeff Haun for providing statistical guidance and timely motivational prompts. My parents certainly deserve my dearest thanks for the years of unwavering support through my academic endeavors. Lastly, I thank my patient husband and beautiful daughter who can now have a wife and mother back!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT ........................................................................................................................ II</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS ........................................................................................................ IV</td>
</tr>
<tr>
<td>TABLE OF CONTENTS ...................................................................................................... V</td>
</tr>
<tr>
<td>LIST OF TABLES ........................................................................................................ VII</td>
</tr>
<tr>
<td>INTRODUCTION ............................................................................................................. 1</td>
</tr>
<tr>
<td>REVIEW OF THE LITERATURE ....................................................................................... 3</td>
</tr>
<tr>
<td>Risk Assessment .......................................................................................................... 3</td>
</tr>
<tr>
<td>Actuarial Risk Assessment ......................................................................................... 4</td>
</tr>
<tr>
<td>Clinical Judgment ...................................................................................................... 6</td>
</tr>
<tr>
<td>Risk Factors for Sexual Reoffense ........................................................................... 7</td>
</tr>
<tr>
<td>Overview of Actuarial Risk Assessment Measures .................................................. 9</td>
</tr>
<tr>
<td>Development of the Static-99 ................................................................................... 11</td>
</tr>
<tr>
<td>Sexual Arousal ........................................................................................................... 17</td>
</tr>
<tr>
<td>Penile Plethysmograph ............................................................................................ 17</td>
</tr>
<tr>
<td>Gaps in the Research ................................................................................................. 22</td>
</tr>
<tr>
<td>PRESENT STUDY ........................................................................................................ 28</td>
</tr>
<tr>
<td>METHOD .................................................................................................................... 31</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Logistic Regression Coefficients – Conviction for any new offense, Static-99 Total Score........................................................................................................36

Table 2: Logistic Regression Coefficients – Conviction for a new sexual offense, Static-99 Total Score........................................................................................................36

Table 3: Logistic Regression Coefficients – Conviction for any new offense, PPG Results..................................................................................................................37

Table 4: Logistic Regression Coefficients – Conviction for a new sexual offense, PPG Results..................................................................................................................37

Table 5: Model Summary – Conviction for any new offense.................................................38

Table 6: Coefficients for Final Model – Conviction for any new offense.............................39

Table 7: Model Summary – Conviction for a new sexual offense...........................................39

Table 8: Coefficients for Final Model – Conviction for a new sexual offense......................39

Table 9: Independent Samples t-Test – Conviction for a new sexual offense...............40

Table 10: Independent Samples t-Test – Conviction for any new offense.........................40
INTRODUCTION

There has been a heightened public awareness of sexual offenders in the criminal justice system in the past 10 years (Melton, Petrila, Poythress, & Slobogin, 2007). Mental health professionals must be able to appropriately address the assessment, treatment, and management of this unique offender group in order to help courts in the endeavor to protect the public. Court officials rely upon accurate assessments of an offender’s future risk for sentencing and release decisions (Weiner, 2006).

Media stories of sex offenders committing new offenses are disturbing to the public, especially if it seems as though the individual should not have been free to have the opportunity to reoffend (Hanson, 2000). The public’s fear of recidivism has been a driving force behind policy makers’ efforts to enact new laws to reduce recidivism. Recent sexual predator laws require civil commitment for sex offenders who are deemed to be too dangerous to release after the offender’s criminal sanction is fulfilled (Doren & Epperson, 2001). As these new laws are enacted to protect the public from predators, psychologists are called upon to relay reliable information about future risk of violent behavior to the courts. Part of these new laws requires mental health professionals to address whether the offender is at an elevated risk for perpetrating future violent acts (Melton et al., 2007; Quinsey, Harris, Rice, & Cormier, 2006).

These recent policy changes raise new ethical and practical questions for psychologists about how to predict risk. In a delicate balancing act between individual
rights and community safety, a professional’s risk assessment can be used to determine if an offender’s civil liberties will be denied and the extent to which the public needs to be protected from the individual. Sexually Violent Predator (SVP) legislation is in effect in at least 16 states and psychologists are routinely asked to conduct risk assessment evaluations to aid courts in determining if an individual meets SVP criteria (Melton et al., 2007). The classification and prediction of sexual offenders have far-reaching consequences, and therefore assessment measures must be as accurate as possible to protect both the offender’s rights and to protect the public.

In this paper, I address the current state of risk assessment with sexual offenders. Data regarding risk assessment with sexual offenders must be continually updated to address the unique needs of the population to ensure that the most accurate methods and measures are employed. The available literature is lacking in information regarding appropriate risk assessment measures for community settings. Additionally, there are surprisingly few studies outlining how to most appropriately interpret and use sexual arousal data provided by the penile plethysmograph. In the current study, I directly address these two concerns by providing information regarding the validity of using the Static-99 (Hanson & Thornton, 2000), an actuarial risk assessment tool, in a community setting as well as exploring the merits and appropriate way to interpret the penile plethysmograph as an objective measure of sexual arousal. To begin with, an overview of risk assessment is warranted as well as a summary of the more specific measures used with sexual offenders. To address the question of whether an offender is at an elevated risk for future violent or criminal acts, psychologists can use several different types of assessment including clinical judgment, structured clinical judgment, and actuarial tools.
REVIEW OF THE LITERATURE

Risk Assessment

Despite early disparaging reviews of the utility of clinical predictions of violence (Craig, Browne, & Stringer, 2003; Monahan, 1981), the American criminal justice system has systematically increased its reliance on mental health professionals to provide predictions of an offender’s violent behavior. Most notably, in Barefoot v. Estelle (1983) the court ruled that mental health professionals were acceptable and reliable experts to comment on the probability that an offender will reoffend, discounting the American Psychiatric Association’s own assertions that mental health professionals cannot provide reliable expert testimony based on general clinical judgment regarding predictions of dangerousness. Over the years, researchers have investigated ways to enhance the reliability of clinical predictions and the utility of a clinician’s opinion to the court. Specifically, the past 20 years have yielded a host of research involving the creation of various risk assessment techniques and measures to cope with the growing field of sexually violent recidivism prediction (Hanson, 2005).

The three primary methods of assessing risk in sexual offenders are unstructured clinical judgment, structured clinical judgment, and actuarial assessments (Rogers & Jackson, 2005). Mental health professionals have noted that clinical judgment based on a psychologist’s professional opinion or “gut feeling” regarding the offender does not produce reliable results (Monahan, 1981). Unstructured clinical judgment is
idiosyncratic, unanchored to empirical research, and results in chance or slightly lower than chance predictions of violence (Monahan, 1981; Quinsey et al., 2006). Some professionals prefer to use structured professional judgment tools that assess variables empirically linked to sexual recidivism and require the clinician to make judgments based on that information. This method yields more reliable results than using only clinical judgment to assess an offender’s risk of reoffense, including sexual reoffense (Litwack, 2001). Other mental health professionals have abandoned clinical judgment altogether, citing the superior statistical performance of actuarial methods that utilize statistical and research guided variables and components (Monahan, 1981; Quinsey et al., 2006). Research has suggested that unstructured clinical judgment carries a weak correlation with recidivism ($r = .08$) in contrast to the typically strong correlation coefficients achieved by actuarial methods ($r = .22$; Craig, Browne, & Stringer, 2003). Reviewing all three types of risk assessment used to assess sexual offenders is beyond the scope of this project; however, a brief overview of actuarial assessment is needed to add context to the current research.

**Actuarial Risk Assessment**

Actuarial risk assessment items are derived from statistically identified correlates of recidivism (Quinsey et al., 2006). Actuarial assessments have largely included historical, static variables that measure unchangeable factors such as demographics or criminal history, yielding an unchanging score, or at least never decreasing score, over time. Static variables are usually easy to quantify and collect from a file, and thus they are easy to include in a statistically derived risk measure with readily available validity information. Though some clinicians prefer clinical methods that give a professional
more flexibility to assess the individual and specific contextual information, Harris et al. (2006) suggested that clinical judgments were subject to human error and not useful as means of conducting a risk assessment. These authors asserted that actuarial methods of risk assessment are far superior to any kind of clinical judgment and that actuarial measures should not be diluted by modifying them with clinical judgment. Actuarial methods are typically well researched and have historically performed well in studies examining the predictive qualities of these measures (Quinsey et al., 2006).

However, actuarial instruments are not without their limitations. First, these methods often do not take into account base rate information (Doren & Epperson, 2001). Substantially lower or higher base rates signify a greater risk for error when trying to predict a specific behavior. Violent acts and sexual recidivism usually have low base rates even in at-risk populations, unless there is a very long follow-up period. Typically, low base rates of violence and sexual deviancy result in a greater number of false positives, or offenders identified as at high risk for reoffending but who do not reoffend. Second, actuarial instruments do not account for clinical change with treatment (Hanson & Bussiere, 1998). Hanson and Thornton (2000), authors of the Static-99 actuarial risk assessment, noted that current actuarial methods exclude the growing literature on the importance of dynamic risk factors that fluctuate over time, such as treatment and alcohol abuse, but did not include such variables in their measure to date. Additionally, applying group level data to specific individuals in making any sort of prediction becomes difficult as the data cannot take into account individual differences or mediating factors as can clinical judgment (Hart, 2003).
Clinical Judgment

In contrast to actuarial risk assessments that have largely used static variables that measure unchangeable factors such as demographics or criminal history, yielding an unchanging score over time, clinical methods give precedent to the professional’s judgment of how to assess the individual and what factors to give more or less weight (Litwack, 2001). As reviewed before, making decisions regarding risk based on clinical intuition or other variables uncorrelated with violence and recidivism has been shown to result in prediction accuracy at chance or less, and some argue that this constitutes poor practice (Monahan, 1981; Quinsey et al., 2006). However, using an assessment strategy that provides a framework for data collection and interviewing, or structured clinical judgment, a clinician is able to evaluate dynamic risk factors that fluctuate and change with life events which a clinician using actuarial assessments may miss by including only static variables. For example, an offender may be considered low risk by actuarial assessment, but if he loses his job and subsequently his housing, he may turn to substance use as a coping skill. All of these stressors are empirically linked with an increase in risk. Structured professional judgment attempts to integrate the static and dynamic factors to allow clinicians to account for the fact that risk level varies across time (Douglas & Skeem, 2005).

There are flexible assessment tools that incorporate a structured interview format to assist clinicians in assessing an offender’s risk to reoffend sexually. Encouraging systematic data collection for relevant risk factors during the interview, structured professional judgment assessments such as the Sexual Violence Risk assessment (SVR-20; Boer, Hart, Kropp, & Webster, 1997) have shown promising predictive qualities
when measuring risk for future sexual violence. The SVR-20 succeeds in producing reliable predictions of future sexual violence (AUC = .83) and relies on gathering historical variables as well as clinical data (Vogel, Ruiter, Deek, & Mead, 2004). However, these tools require an investment of the clinician’s time over and above actuarial assessments, and validation literature is lacking due to the relative novelty of this approach to risk assessment. Though clinical experience should not be overlooked, an unstructured assessment left entirely up to the individual clinician opens many issues of professional liability for clinician both in the community and in the courtroom.

Clinician’s decisions should be supported and guided by research. Currently there is convincing evidence for the use of actuarial assessments, although not to the exclusion of other forms of clinical assessments because actuarial methods are not without controversy or error in predicting recidivism. Which assessment methods provide the most accurate results in the area of risk assessments with sexual offenders is unknown at this point. However, it is clear that actuarial assessments are statistically superior to unstructured clinical judgment at predicting violent and sexual recidivism (Quinsey et al., 2006).

*Risk Factors for Sexual Reoffense*

The last 20 years have yielded significant improvements in understanding risk for sexual recidivism with improved statistical methods and evolution of enhanced types of analysis, such as the meta-analytic approach. Combining multiple study samples into a meta-analysis allows researchers to view the scope of the current findings and increase statistical power; however, the investigator must also reconcile differing methods and definitions across studies (Hanson, 2005). Another advancement is the use of Receiver
Operating Characteristic (ROC) curves, and corresponding Area Under the Curve (AUC) statistics, to interpret experimental results (Swets, 1986). AUC statistics are unaffected by base rates, and thus are particularly useful in research with sex offender populations where offense rates are estimates at best and typically low. In interpreting AUC output, .50 equals predictive accuracy of 50%, or chance level (Swets, 1986; Zou, O'Malley, & Mauri, 2007). Numbers above .50 representing predictive quality beyond chance, and the AUC can be tested for whether it is statistically significant in difference from chance. AUC values of .70 are generally accepted as indicating good predictive ability (Hanson & Thornton, 2000; Zou et al., 2007). These two methods of analyzing results have helped produce meaningful gains in the knowledge of sexual offence recidivism risk factors.

Results of recent meta-analyses have revealed factors highly correlated with sexual offending to be deviant sexual arousal, general criminal history, and psychopathic personality (Beech, Fisher, & Thornton, 2003). Hanson and Bussiere’s (1998) meta-analysis of risk factors of sexual recidivism remains one of the most well-regarded studies to date and reflects these findings. Included in their analysis were only studies in which researchers assessed a group of sex offenders for characteristics and correlated those factors with recidivism after a specified follow-up period. A total of 61 studies with unique data sets were collected, resulting in a sample size of 28,972 sexual offenders. Using correlations to measure predictive accuracy of each criterion, the authors found the greatest predictors of sexual recidivism to be deviant sexual arousal as measured by phallometry ($r = .32$) and a history of prior sexual crimes ($r = .19$). To a lesser degree, young age ($r = -.13$), presence of a personality disorder ($r = .13$) or psychopathy ($r = .14$), and never-married status ($r = .11$) were also correlated with sexual recidivism.
Psychopathy as measured by the Psychopathy Checklist, Revised (PCL-R; Hare, 1991) has been highly correlated to recidivism in general and is often included as an item to be scored in risk assessment measures. However, the PCL-R has been shown to be less predictive of sexual recidivism than general or nonsexually violent recidivism at a statistically significant level (Barbaree, Seto, Langton, & Peacock, 2001).

Hanson and Morton-Bourgon (2005) updated Hanson and Bussiere’s (1998) study and found similar results. The authors examined 82 recidivism studies and reduced specific variables into seven categories: sexual deviancy, antisocial orientation, sexual attitudes, intimacy deficits, adverse childhood environment, general psychological problems, and clinical presentation. The authors also found deviant sexual arousal to be the strongest predictor of sexual recidivism ($r = .30$) followed by antisocial orientation ($r = .23$). Among other things, these two studies revealed the importance of including a measure of deviant sexual arousal when assessing risk for sexual recidivism. Interestingly, these results suggest that some of the more common treatment targets when working with sexual offenders, such as intimacy deficits, accepting responsibility, and cognitive distortions regarding sexual offending (Hall, Shondrick, & Hirschman, 1993), were not significantly related to recidivism. Also important to note, deviant sexual arousal and antisocial orientation arguably may be considered dynamic risk factors that may be targeted for change during treatment. Armed with those correlational data about risk factors, clinicians and researchers have been able to craft multiple measures and tools to assess and predict an offender’s risk for reoffense to varying degrees of success.

Overview of Actuarial Risk Assessment Measures

Clinicians have been able to combine the risk factor research to create items for
actuarial tools that assess risk. The Violence Risk Appraisal Guide (VRAG) is an actuarial measure that has demonstrated an ability to predict any form of recidivism and that was a pioneering measure in both empirical construction and predictive qualities (Quinsey et al., 2006). The VRAG consolidated known predictors of violent recidivism into one measure. Items that showed predictive power included static variables from childhood home, elementary maladjustment, age, criminal history, victim characteristics, prior release failure, psychological disorders, marital status, and Psychopathy Checklist, Revised (PCL-R) score. These variables were found to be significant predictors of violent recidivism based on a sample of male Canadian forensic mental health patients with a correlation between the items and reoffending of at least .11 (Harris et al., 2003).

The validity of the VRAG has been established for predicting recidivism for violent offenders as well as sex offenders in multiple studies (Quinsey et al., 2006). Rice and Harris (1997) attempted to validate the VRAG with a group of 288 male sex offenders in Canada. The authors showed the predictive qualities of the VRAG with an AUC of .60 for sexual reoffending and .76 for violent reoffending after a 10-year follow-up period. This result was mirrored in a study comparing six risk assessment measures in predicting recidivism with a shorter follow-up time of 4.5 years in which the VRAG outperformed the other measures (Barbaree et al., 2001). With 215 participants, the authors found that the VRAG performed well with an AUC of .61 for sexual recidivism and .77 for any form of recidivism when examining recidivism with sexual offenders.

Created by the same authors as the VRAG, the Sex Offender Risk Appraisal Guide (SORAG) built on the success of the previous measure and, despite its moniker, is used to determine the level of risk for any offender to reoffend sexually or violently with
purportedly more predictive validity (Quinsey et al., 2006). Two items were added to the SORAG to make it more sensitive for sexual offenders including previous sexual crime history and an item reporting deviant sexual arousal. The SORAG displays similar AUC values as the VRAG at .70 for any recidivism and more specificity than the VRAG for sexual reoffense at .70 (Barbaree et al., 2001).

Another newer actuarial risk assessment measure that was developed separately from the VRAG and that has shown promise at predicting sexual and violent recidivism is the Static-99 (Hanson & Thornton, 2000). The VRAG and SORAG have demonstrated excellent predictive power when predicting sexual and violent recidivism with sexual offenders; however, they have also proven to be cumbersome when used in a community setting (Hanson & Thornton, 2000). Though many offenders are sent from a prison or institution to treatment in the community, the treatment facility may not have access to records or resources to conduct the PCL-R on each offender. These obstacles render the impressive validation data less useful when the data for the full assessment are not available and the predictive accuracy is reduced. Thus, more practical, but still empirically sound, actuarial assessments, such as the Static-99, must also be available for clinicians.

Development of the Static-99

Hanson and Thornton (2000) created the Static-99 as a measure that consolidated two shorter risk assessments, the Structured Anchored Clinical Judgment (SACJ; Grubin, 1998) and the Rapid Risk Assessment for Sex Offence Recidivism (RRASOR; Hanson, 1997), into a single measure. The RRASOR was developed to provide an easily scored and quick assessment of an offender’s risk to reoffend. Pulled from Hanson and
Bussiere’s (1998) meta-analysis as the most correlated factors to recidivism, the RRASOR consists of four items: number of prior sex offenses, offender’s relation to the victim, gender of the victim, and the offender’s age (Hanson & Thornton, 2000). The RRASOR was found to have acceptable predictive validity ($AUC = .71$) in a validation study using an existing database consisting of sexual offenders with a follow-up time of 4 to 23 years (Hanson & Thornton, 2000). The authors of the SACJ attempted to predict both sexual and violent recidivism using an adjustable level approach (Grubin, 1998; Hanson & Thornton, 2000). To date, there has been little research demonstrating the empirical value of the SACJ. Using an exploratory analysis approach, the authors of the SACJ used existing data from the RRASOR validation study of incarcerated individuals in the United Kingdom. The initial level consists of the offender’s criminal record and the second level evaluates aggravating factors, such as relation to the victim. The final level attempts to take into consideration dynamic variables such as treatment outcomes.

Hanson and Thornton (2000) used the RRASOR and the criminal history variables of the SACJ to predict sex offender recidivism and suggested that these measures were tapping into similar yet separate constructs because the measures contributed unique variance in preliminary regression equations, though the authors did not define the contents or boundaries of those distinct concepts (Hanson & Thornton, 2000). When combined, Thornton and Hanson (2000) attempted to provide a more complete risk assessment of male sex offenders and named it the Static-99. The new measure contained static risk factors such as demographics and offense details with the goal of measuring risk for sexual recidivism. It contains 10 items and, as the name suggests, all items are static variables. The Static-99 is scored using a preset weighted
score for each item similar to the VRAG, with a higher score indicating greater risk of future violence. The total score is calculated by simple addition of the items with a minimum score of 0 and a maximum score of 12. The offender is then assigned one of four risk categories: A score of 0 to 1 corresponds to low risk, a score of 2 to 3 is medium-low risk, a score of 4 to 5 is medium-high risk, and a score of 6 and over is considered high risk.

The original validation study by Hanson and Thornton (2000) held promising results for the Static-99. The validation sample consisted of four groups of male sexual offenders from different facilities in the Canadian criminal justice system at different points of time for a total of 1,208 participants. These offenders were detained in either a prison or a psychiatric facility and the researchers maintained a follow-up period of 4 to 23 years. The authors reported the predictive accuracy of the Static-99 as significantly above chance for sexual recidivism ($AUC = .71$) and violent recidivism ($AUC = .69$). As predicted, these scores were also higher than the RRASOR (.68 for sexual recidivism, .64 for violent recidivism) and the SACJ (.67 for sexual recidivism, .64 for violent recidivism).

In an archival study with 396 male sex offenders, Harris et al. (2003) assessed the predictive accuracy of the VRAG, SORAG, RRASOR, and Static-99. The Static-99 ($AUC = .63$) performed better than the RRASOR ($AUC = .56$) but not as well as the VRAG ($AUC = .73$) when predicting violent recidivism. When comparing performance of the measures on predicting sexual recidivism, the Static-99 performed well with an AUC of .65 compared to the VRAG and SORAG which each obtained an AUC of .70. The result was promising for the Static-99 as a specific tool to assess sexually violent
offenders, though it did not perform as well as the VRAG and SORAG. Interestingly, Harris et al. found that deviant sexual arousal showed lower correlation with sexual recidivism \((r = .09)\) than had been reported in more recent studies (Hanson & Bussiere, 2005), thus indicating some disagreement in the field regarding the degree to which deviant sexual arousal predicts sexual reoffense rates. However, it should be noted that the Harris et al. finding is from one study, whereas the Hanson and Mouton-Bourgon (2005) finding is a result of a meta-analysis of eight studies.

In a smaller study, Duero and Pham (2006) directly compared the Static-99 and the SORAG with a sample of 147 Belgian male sexual offender participants who were confined to a psychiatric hospital and followed for an average of 4.2 years after release. The study was a retrospective design using patient file data and many items were missing for the scoring of the SORAG, such as the PPG (100% missing) and the PCL-R (66% missing). Though not statistically significant, the Static-99 performed better than the SORAG in predicting sexual recidivism with AUC values of .66 and .64 respectively, and performed equally well in predicting general recidivism (AUC = .70). However promising these results appear for the Static-99, the data seem to support the importance of the PPG and PCL-R items for the SORAG rather than supporting the superior performance of the Static-99. On the other hand, the Static-99 does not require items that are more difficult to collect and score, such as the PCL-R and childhood adjustment information.

Dorian (2004) attempted to validate the risk categories of the Static-99. Using data sets from 12 previous research ventures studying the recidivism of sex offenders, the author determined that the Static-99 risk categories of low, medium-low, medium-high,
and high were stable across all the studies despite the varying base rates of recidivism.

Dorian examined the recidivism rates in each risk category and compared those rates across the studies with varying base rates. This study was noteworthy given the difficulty of estimating population-wide base rates of sexual recidivism.

Researchers studying the Static-99 have produced a small but consistent and promising body of literature demonstrating its validity. However, many of these studies consist of samples drawn from confined inpatient or prison populations (Hanson & Thornton, 2000), thereby limiting external validity when extending the measure beyond prison walls to a community setting. In other words, these measures have been validated and studied with offenders who have been incarcerated, but not with individuals who have not been incarcerated for their sexual deviancy. Thus, there is a question about the utility of this measure with individuals who have not yet been incarcerated. Also, most of the data used were from an existing database also used in the validation studies of other actuarial measures, such as the VRAG and RRASOR (Hanson, 1997). Examining the utility of the Static-99 with more heterogeneous samples is required in order to extend clinical relevance of risk assessment in a community sample. Having a quick and easy measure with comparable validity coefficients relative to more time-consuming methods makes the measure more useful on a wider scale if validity can be well established.

Recent attempts to validate the use of the Static-99 with samples representing differing custody levels and geographic locations have led to a confusing, controversial, and conflicting body of literature (Helmus, Hanson, & Thornton, 2009). Promising findings in initial validation studies gave way to less impressive results that fell below best standards of risk prediction (Hart, Michie, & Cooke, 2007; Helmus et al., 2009). In a
recent meta-analysis, Hanson and Morton-Bourgon (2009) examined 118 studies exploring the prediction of recidivism among sexual offenders and found that the Static-99 performed at chance level when predicting both any recidivism and violent recidivism ($AUC = .51$ and .53, respectively), but it did attain an acceptable predictive value when predicting sexual recidivism ($AUC = .67$). Likewise, in another recent researcher’s study found that the Static-99 performed poorly when predicting sexual violence ($AUC = .55$), general sexual recidivism ($AUC = .57$), and non sexual violence ($AUC = .49$) in a sample of offenders released from a penitentiary in the United States (Boccaccini, Murrie, Caperton, & Hawes, 2009). To manage new information regarding the poor performance of the Static-99, the Static-99R was introduced with the addition of a refined age item and a possibility to obtain a negative score (Harris et al., 2008). Additionally, the new norms included instructions for using clinical judgment to fit an offender into “preselected” risk groups. However, many clinicians are resistant to using the new norms due to the lack of completed research and a dearth of information regarding the definitions and inclusion criteria for the preselected groups (Helmus, 2007). As the controversies are sorted out through further research and proper use of the measure is clarified, continued use of the Static-99 remains clinically acceptable and widely used in the sexual risk assessment field (Hanson & Morton-Bourgon, 2009; Helmus, 2007).

Improvement on risk assessment methods should always be explored given that no prediction scheme accurately predicts 100% of the time. Like the VRAG, the Static-99 may benefit from adding items to improve the predictive qualities of the measure. For example, as it currently stands, the Static-99 does not include a measure of deviant sexual arousal, which is an important dynamic risk factor as previously noted (Hanson, 2005).
The addition of a measure of deviant sexual arousal may indeed improve this already promising measure.

**Sexual Arousal**

Although there are many psychological and physical components to offending, sexual arousal appears to be one of the key elements to a crime of sexual deviancy. There is a strong relationship between sexual arousal and an offender’s propensity to reoffend sexually (Hanson & Bussiere, 1998). Researchers studying the relationship between deviant sexual arousal and recidivism have found correlations of .19, placing deviant sexual arousal as one of the most predictive variables of sexual reoffense, well above the next most predictive variable of criminal lifestyle ($r = .12$; Hanson & Bussiere, 1998).

Perhaps the ultimate actuarial item to predict reoffense with sexual offenders would be an objective measure of the degree to which an offender becomes aroused to deviant sexual stimuli. Such a measure would provide the risk assessor with strong evidence of the offender’s risk to reoffend, though certainly the information should be used in conjunction with other pieces of evidence to provide a more complete picture of an offender’s risk to reoffend. Currently, the most widely excepted measure of deviant sexual arousal is the penile plethysmograph (Tong, 2007).

**Penile Plethysmograph**

Phallometry, or specifically the use of penile plethysmograph (PPG), has often been touted as the most reliable and commonly used form of measuring deviant sexual arousal in sex offenders (Able, Huffman, Warberg, & Holland, 1998; Barker & Howell, 1992; Tong, 2007). A flexible ring is placed at the base of the penis, which measures the change in circumference, or tumescence, of the penis due to blood flow in response to
arousing stimuli. The kind of stimulus used is often up to the administrator of the measure, but the two most common forms are audio recordings of verbal scripts detailing sexual or erotic acts and visual slides depicting clothed and unclothed children and adults. The verbal stimulus method uses a tape-recorded message describing people and situations and measures an offender’s reaction to each section of the script. Visual stimulus in a PPG evaluation uses images of people and situations and measures the offender’s reaction to each stimulus (Simon & Schouten, 1991).

PPG is used by mental health professionals to assess sex offenders with the assumption that there is a link between the PPG arousal in response to specific stimuli and behavioral acts of sexual aggression in the community. As previously discussed, deviant sexual arousal as measured by the PPG is a significant risk factor for general recidivism and, when used properly, it can aid in providing an accurate reflection of an offender’s risk to reoffend sexually (Hanson & Bussiere, 2003).

Beginning in 1963 with Freund’s influential research on reliable measurement of sexual arousal with phallometry, there have been many studies touting the basic reliability and validity of the PPG (Freund, 1963; Launay, 1994). Researchers have consistently found statically significant differences between the arousal patterns of groups of participants with a history of deviant sexual acts and control groups of individuals without histories of sexual offenses (Launay, 1994). Recently, Able et al. (1998) conducted a comprehensive study using the PPG to measure the sexual arousal patterns of 157 male participants referred to an outpatient clinic treating deviant sexual behavior. All participants were volunteers and had admitted to inappropriate sexual behavior. The researchers compared each participant’s physiological response to visual
slides as measured by the PPG to visual reaction time, another purported measure of sexual arousal. The authors reported the visual slides procedure for the PPG to be reliable with a Cronbach’s alpha coefficient between .66 and .97 depending on the age and gender depicted on the slide. Measurement validity was also established for the PPG as the authors found statistically significant correlations between self reported arousal as well as reliable categorization of sex offender typologies with PPG results.

The stimulus set is also an important consideration when discussing the validity of the PPG. Harris et al. (1992) found both auditory and visual stimulus sets for PPG measurement to be valid and reliable. These authors also offered suggestions for enhancing phallometric testing reliability and validity, including using the most sexually vivid stimulus set and using z-scores or a standardized unit of measure to communicate relative arousal unique to the individual. Malcolm, Andrews, and Quinsey (1993) also tested a variety of stimuli in their evaluation of the discriminant and predictive validity of the PPG using 172 incarcerated male sexual offenders. The authors found that offenders showed statistically significant arousal to stimuli depicting the gender of the victim corresponding with the offender’s index offense. Offenders who recidivated also showed statistically significant arousal to deviant age preference stimuli, an effect that was magnified when applied to participants convicted of sexual crimes against children. The authors also found that they could reliably predict recidivism as it related to deviant sexual arousal to stimuli depicting underage individuals, but they could not reliably differentiate familial from non-familial child molesters. These findings represent a consistency in the literature showing that PPG results can provide helpful and reliable physiological information to aid professionals in predicting recidivism.
Though phallometry may be the best tool available for measuring deviant sexual arousal, it is not without controversy. Researchers have been able to reliably parse non-offender from sexual offender PPG profiles (Howes, 1995); however, there have been warnings from researchers against classifying preferences within offender groups because conflicting results have arisen when attempting to differentiate categories of sexual offenses or identification of sexual orientation (Launay, 1994). That is, research has supported the use of the PPG to determine whether an offender demonstrates deviant sexual arousal but not necessarily to say whether the offender is attracted to a particular gender or victim type. More importantly, researchers have strongly cautioned against determining guilt of an offender based on PPG results; there have been no studies suggesting that this would be an ethical or empirically validated use of the data (Howes, 1995; Launay, 1994; Fernandez, 2009). Many researchers have also contended that the results of phallometric testing can be manipulated, thereby rendering invalid results (Howes, 1995; Launay, 1994). This criticism has received some support in that researchers have shown that individuals can decrease the degree of arousal to certain stimuli, especially under repeated testing such as is required in some sex offender treatment protocols (Harris et al., 1992; Lalumiere & Harris, 1998). The strength of these criticisms, however, is lessened when only an initial PPG is required to determine deviant from non-deviant sexual arousal patterns. However, the need remains for developing the best possible technique to use the data gained from PPG protocols.

Often the specific stimulus method for the PPG protocol are left up to the evaluator given that research has shown similar validity for visual and verbal stimuli
The external validity is called into question due to the variety of administration methods and the imposition of a laboratory environment on arousal, and many researchers have strongly cautioned against making decisions of guilt based on the results of the PPG (Able et al., 1998; Fernandez, 2009). The PPG is typically limited to measuring deviant sexual arousal in adult males with a history of sexual aggression (Able et al., 1998; Simon & Schouten, 1991). The Association for the Treatment of Sexual Abusers (ATSA) established guidelines on the use of and equipment for PPG evaluations but fell short of specifying the stimulus set or scoring method that should be used (ATSA, 2004). There are guidelines to combat malingering and poor testing methods, such as using a continuous measure to obtain a baseline and using the most explicit stimuli available (Barker & Howell, 1992; Harris et al., 1992); however, these guidelines are not mandates, and not all evaluators use the solid methods purposed by these authors, bringing into question the generalizability of the highly controlled research results to a community setting.

Unlike the development of actuarial risk assessment measures, the past 20 years of study into plethysmographic technique and interpretation have not paralleled the growing need for precision and accuracy in the risk assessment field (Howe, 1995). Procedures and interpretation methods have remained virtually unchanged despite growing conflict among experts. At the same time, courts demand the highest level of predictive accuracy possible from mental health professionals in order to correctly classify and sentence offenders (Melton et al., 2007). Adding to the difficulty in establishing the PPG as a risk assessment tool, research results have become outdated and the studies were performed in highly controlled settings with confined individuals, not in
a community setting. One of the most troublesome criticisms of the PPG is the lack of set interpretation standards in the field. Recent research has highlighted that the threshold labeling an offender as possessing deviant sexual arousal based on PPG results alone varies widely from 10% to 30% tumescence (Fernandez, 2009). The variability between sites and administrations of the PPG hampers efforts to conduct research because an offender deemed as having deviant sexual arousal by one standard will not match up with another offender in a different location or research protocol. Additionally, the weighty determination of deeming an individual at risk to sexually reoffend is often not anchored in empirically derived values and is instead based on an individual PPG administrator’s clinical interpretation (Howes, 1995).

Despite the controversy and lingering questions, the PPG has demonstrated reliability and validity that exceeds many commonly used psychological measures (Able et al., 1998; Tong, 2007) and can aid professionals when assessing sexual offenders. Simply stated, PPG provides a physiological measurement of deviant sexual arousal, the most highly correlated risk factor for sexual recidivism. Given the ambiguity of conflicting research results in the literature, the most appropriate way to interpret and report PPG results remains under discussion. PPG results are currently generally interpreted in a dichotomous manner in which the evaluator decides if the offender shows sufficient arousal to deviant sexual stimuli to indicate elevated risk (Fernandez, 2009). However, it may prove fruitful to explore and validate a continuous scale representing more or less risk of reoffense depending on arousal patterns.

Gaps in the Research

In a recent pilot study, the utility of the Static-99 and PPG results in a community
setting was called into question (Means, 2008). Contrary to previous studies (Able et al., 1998; Hanson & Thornton, 2000), neither the Static-99 total score nor the presence of deviant sexual arousal as measured by PPG was predictive of sexual recidivism for male sexual offenders presenting for treatment for sexual deviancy in a community treatment facility. Furthermore, PPG results did not add incremental validity to the Static-99 when predicting sexual recidivism, as was expected given the previous use of deviant sexual arousal as an item on the VRAG (Quinsey et al., 2006). Using AUC statistics to measure the predictive quality of the Static-99 in the pilot study sample, there was a trend indicating that the Static-99 was able to predict conviction for a new sexual offense to a similar degree as has been found in other studies (Hanson & Thornton, 2000; Means, 2008), though not at a statistically significant level. These results may have been due to the lack of statistical power and low rates of conviction for a new sexual offense (10% of sample), both of which were affected by the small sample size in that study ($N = 50$). In addition, many risk assessment measures can perform well in highly controlled validation studies and the results may not translate to community settings. Few validation studies on risk assessment measures have been completed with non incarcerated participants to date. Given the non-significant results of the pilot study, larger scale studies need to be conducted to shed light on the conflicting evidence of using the Static-99 and PPG to predict recidivism with sexual offenders.

Assessment of sexual offenders may benefit from the inclusion of a measure of deviant sexual arousal given that this factor has been shown to be highly predictive of sexual recidivism (Hanson & Morton-Bourgon, 2005). PPG is the best objective measure of deviant sexual arousal currently available (Barker & Howell, 1992; Lalumiere &
Harris, 1998) and its utility in the evaluation and treatment of sex offenders has been demonstrated. Though PPG measurement is not without controversy, there are few other options for psychologists to use to determine deviant sexual arousal. Deviant sexual arousal was added to the SORAG, improving the predictive qualities of the original VRAG from an AUC value of .60 to for sexual recidivism to .70 for the SORAG (Quinsey et al., 2006). These results showed, in part, the value of the PPG in increasing the accuracy of risk assessment measures. The literature reviewing risk assessment with sexual offenders outlines the importance of objectively measuring sexual thoughts and preferences towards children or harm to others (Haywood, Grossman, & Cavanaugh, 1990; Tong, 2007). The PPG with the appropriate stimulus set can determine an offender’s propensity towards deviant sexual arousal, and that information may assist in conducting a thorough risk assessment.

Though there have been calls to include more fluid dynamic factors that have been empirically identified as risk factors for sexual recidivism (Douglas & Skeem, 2005; Hanson & Thornton, 2000), this research remains in its infancy. Dynamic risk factors capture a current picture of an offender’s risk, including factors such as deviant sexual arousal, substance abuse, and lifestyle changes (Haywood et al., 1990). Such factors, which fluctuate over time, create a challenge for psychologists’ ability to develop a measure able to capture an offender’s risk with a one-time assessment tool without losing important information about that offender’s risk. Some researchers are beginning to add dynamic factors into actuarial risk assessment measures (Hanson & Thornton, 2000), but progress remains slow. As an example, deviant sexual arousal was added to the SORAG to increase predictive accuracy for sexual recidivism from an AUC of .60 for the VRAG
to .70 for the SORAG (Quinsey et al., 2006). However, the scale is currently recommended for use as a predictor of general violent recidivism, not specifically for sexual recidivism.

Though the PPG may be the best available tool for measuring deviant sexual arousal, there is little research regarding the appropriate way to interpret individual results. Not only were very few peer-reviewed articles regarding the PPG available to review in preparation for the current study, authors of previous research also rarely included operationalized definitions for measuring deviant sexual arousal. In a survey of clinicians using the plethysmograph in practice, an alarming degree of variance was reported in interpreting the results (Howes, 1995). The author found that clinicians interpreted the cut-off for clinically significant arousal to deviant stimulus between 5% and 20% of the offender’s full erection. The variance in interpreting clinically significant arousal seems to represent an ongoing struggle to maintain clinical autonomy while setting industry-wide standards on how to appropriately interpret PPG results. Given the lack of consensus in the field, perhaps a fresh perspective on interpretation of PPG data would be beneficial. A dimensional model of PPG scores has not been thoroughly researched but may be a more accurate way of reflecting an offender’s risk to reoffend based on PPG results.

Very few researchers have outlined the threshold used to interpret deviant sexual arousal from the data obtained in a PPG protocol. Perhaps due to the lack of reliable research to guide decisions, here are virtually no firm standards of practice as to the specific threshold of deviant and non deviant arousal as measured by the PPG set forth by professional associations. The most common method is a dichotomous description of
arousal, reducing widely variable results to a simple yes or no check mark. The prior pilot study (Means, 2008) examining the predictive qualities of the Static-99 and PPG indicated that there is still work to be done when attempting to predict sexual reoffense with physiological and actuarial risk assessments in community settings. The results showed that a large number of participants displayed deviant sexual arousal by achieving 5% of full arousal, but the majority of those with deviant sexual arousal did not sexually recidivate. Although the PPG appears to be a valid instrument to use with sexual offenders to measure deviant arousal (Able et al., 1998; ATSA, 2004), the threshold standard may need to be revisited to evaluate whether or not a threshold of arousal model is the most appropriate application of the data. A dimensional model of PPG results may provide more accurate prediction of reoffense than a simple presence-absence dichotomous variable. An offender who shows 5% deviant sexual arousal on a stimulus may present a very different risk to reoffend than an individual displaying 90% arousal to the same stimulus. Similarly, an item in a risk assessment reflecting a dimensional model would be a dynamic variable that would fluctuate over time and with treatment. This type of model has yet to be tested and there is little to no literature devoted to the topic of a dimensional model of the PPG measure.

There have been conflicting results when reporting on the predictive qualities of various risk assessment measures, and there is little consensus as to the best measure to use when attempting to conduct risk assessments with sexual offenders. The ambiguity when assessing risk with offenders in the community is especially apparent when dynamic variables such as housing, treatment, and substance abuse are present (Means, 2008; Tong, 2007). As new information arises from the literature regarding predictors of
recidivism, researchers must be prepared to incorporate that knowledge into new or existing measures. The accurate prediction of sexual recidivism is far from perfect, and future researchers should strive towards the most accurate level of prediction in order to ensure protection of the public as well as preserving the rights of the offender. As more programs are developed to cope with the growing multitude of individuals who are registered sex offenders, validation in a community setting is crucial for both the Static-99 and PPG to be used appropriately.
PRESENT STUDY

Given that those labeled “predatory sex offenders” continue to be adjudicated and subsequently civilly committed at a high rate, the growing field of sex offender risk assessment should be supported by current studies and reviews to clarify the utility and limits of risk assessment procedures. Much of the data collected and used to validate current risk measures are from one expansive data set (Hanson & Thornton, 2000; Quinsey et al., 2006), limiting the generalizability and utility of the measures. As the pilot study found, there are conflicting results as to the reliability of risk assessments outside of incarcerated populations (Means, 2008). In the current study, I examined the utility of the Static-99 in predicting conviction for sexual offenses with a community-based sample of men being treated for sexual deviancy. To promote generalizability of the measure, the Static-99 was retrospectively applied to archival data of a large sample of these subjects. My first hypothesis was that the Static-99 would predict conviction for a sexual and general offenses at a rate similar to those noted in the validation samples.

Deviant sexual arousal is a strong predictor of sexual recidivism (Able et al., 1998; Hanson & Bussiere, 1998). Although the utility of the PPG has been called into question because of the invasive nature of the procedure (Tong, 2007), if it is a strong predictor of sexual recidivism, the benefits of an accurate assessment of risk may outweigh the invasion of privacy associated with the procedure. The pilot study indicated that a dichotomous cut-off of 5% arousal to deviant stimuli was ineffective at predicting
new convictions. Thus, in the current study, I used a dimensional model in which the presence of more arousal to deviant stimuli was assumed to incrementally increase the risk of reoffense. In the current study also attempted to find predictive validity in a community sample of sexual offenders to determine the utility of using this dimensional model of the PPG as a predictive tool. My second hypothesis was that a dimensional model of PPG results would adequately predict new convictions for both sexual and general crimes.

As other researchers have shown in incarcerated populations (Hanson & Thornton, 2000), an underlying assumption of the current study was that an increase in the Static-99 total score signified an increased risk of reoffense. Additionally, it was also assumed that the higher the level of arousal to deviant sexual stimuli, the higher the risk to reoffend (Able et al., 1998; Tong, 2007). To test this assumption that both measures are measuring a similar construct of risk to reoffend, my third hypothesis stated that as the total Static-99 score increased, the percent of arousal measured by the PPG would also increase, resulting in a positive correlation between the two measures.

As mental health professionals have previously noted (Hanson & Thornton, 2000), adding dynamic risk factors, specifically the highly correlated factor of deviant sexual arousal, may raise the predictive abilities of existing actuarial assessments. In the present study, I examined the incremental validity of adding a PPG measure to the Static-99. The fourth hypothesis stated that adding the initial PPG score to the Static-99 would add incremental validity to the measure in predicting conviction for sexual and general offenses. Adding this important empirically supported dynamic risk factor to the static, actuarial-based Static-99 suggests increased ability to identify which sex offenders will
In summary, I attempted to add to the literature by investigating the following hypotheses: (a) The Static-99 total score would predict new general and sexual convictions, (b) a dimensional model using percent of arousal would be predictive of new sexual and general convictions, (c) there would be a positive correlation between the Static-99 total score and increased PPG percentage of arousal, and (d) adding a dimensional score of PPG would add incremental validity to the predictive qualities of the Static-99 total score when predicting new convictions for general and sexual offenses. Though the pilot study showed no statistically significant results when examining the predictive qualities of the Static-99 and PPG with a small community sample, the current study is expected to gain statistical significant results due to a number of factors. These factors include; a new perspective for interpreting PPG data and increased statistical power due to a larger sample size and a longer follow-up period.
METHOD

Subjects

An archival data collection method was utilized in the current study. Data were gathered from a sample of 131 male sexual offenders presenting for evaluation and/or therapy at a community sex offender treatment facility in Northwestern Oregon. The average age of subjects was 36 years old (range = 18-68, SD = 12). The most common offenses were child sexual abuse (64%), followed by rape (11%), exposure (7%), sexual abuse of an adult (6%; e.g., sexual contact with a vulnerable adult), and other (12%). The racial composition of the sample was 80% Caucasian, 5% African-American, 5% Latino, 2% Native American, 2% multiracial, and 6% unknown. All participants were under legal jurisdiction (such as pending charges or under supervision) at the time of the evaluation and had been mandated to participate in an assessment to determine sexual deviancy risk.

The primary school educational level of this sample ranged from 6th grade to some college, with 110 (84%) subjects completing high school or graduation equivalency. Twenty-nine percent of the subjects (n= 38) had received vocational training, and 24% attended college or university classes. In regard to career choices, 76 (58%) were laborers, 16 (12%) were professionals (e.g., health care professionals, office professional, teacher), 3 (2%) worked in law enforcement, and 36 (28%) had worked in other settings. Thirty-six subjects (24%) had been in the armed forces. The majority of the sample (63%) were unmarried at the time of the evaluation, but 54% had been
married at some point.

Measures

Records for clients presenting for psychosexual evaluations were examined to obtain PPG evaluation results, demographic data, and information to score the Static-99. Each subject underwent a full psychosexual evaluation upon initiating contact with the treatment facility, encompassing social and sexual history information, a polygraph examination, a clinical interview, personality testing, and a PPG evaluation.

*Penile Plethysmograph (PPG)*

PPG assessments were conducted before treatment at the initial assessment and were administered per the guidelines set out by the Association for Treatment of Sexual Abusers (ATSA). Each client met with a male examiner who introduced the assessment procedures and reviewed the informed consent. The participant entered a private room, sat in a recliner, and followed instructions to place the gauge on his penis. Audiotapes and visual slides were both used on participants. The tapes and slides depict various ages and levels of coercion. Evaluators were trained to detect if the gauge remained on the offender’s penis and was correctly recording data. Additionally the computer program used to record the PPG data was equipped to analyze output patterns and detect malingering or equipment failure. An attention check was also included, which involved the client pressing a button when a prompt was randomly displayed during the evaluation. Clients obtained a baseline arousal to each stimulus that was compared to arousal data obtained for different sexual stimuli and a percent of change in arousal was recorded.

In the present study, I examined the records of clients and recorded the percent of arousal in relation to his risk to reoffend. The highest arousal to any deviant stimuli
demonstrated by each subject was used to analyze the proposed dimensional model of interpreting the PPG.

*Static-99*

As previously discussed, the Static-99 (Hanson, & Thornton, 2000) is comprised of 10 items. A total score is calculated between 0 and 12 by adding scores obtained on each item, with higher scores representing more risk to recidivate sexually or violently. The total scores are divided into four levels of risk based on statistical risk to reoffend: low, medium-low, medium-high, and high. For the purposes of this study, participants’ numerical score was used, rather than the categorical rating of risk level.

**Procedure**

The current study used an archival data set from a community treatment center for sexual offenders. Pacific University Institutional Review Board approval was obtained before proceeding with data collection. Due to a number of research assistants entering data, a coding manual was created to dictate the proper classification and entry of each variable. Each file was reviewed, and pertinent variables were entered into the database. The data set was verified for accuracy with a random 10% of the files checked for erroneous data entry by the chief researcher of the larger project.

One hundred thirty one subjects from the larger database of clients were included in the current study. The files of subjects included in the current study had a valid pre-treatment PPG and enough demographic and psychosocial history to score the Static-99. Subjects from the larger database were not included in this study if there was insufficient information to score the Static-99 or if the PPG results were invalid due to inadequate baseline level of arousal (e.g., arousal responses were low to all stimuli). The highest
PPG reading to deviant stimuli obtained by each participant during the evaluation was used in statistical procedures. The Oregon Department of Corrections provided access to a criminal history database and reconviction data for each participant was collected retroactively from the time the offender entered the treatment program.
RESULTS

Oregon Department of Corrections data indicate that 20 subjects (15%) recidivated in any way after a mean follow-up time of 8.82 years ($SD = 5.59$, range = 2-21). Of those who recidivated, 13 subjects (10%) were convicted of a new sexual crime and 10 subjects (8%) recidivated with general charges, such as probation violations or driving while intoxicated. No subjects in the current sample were convicted of a nonsexual violent crime.

The mean score across all subjects on the Static-99 was 2.47 ($SD = 1.74$, range = 0-7), indicating a low risk to reoffend (Hanson & Thornton, 2005). PPG results indicated that the mean highest percentage obtained by subjects was 47% tumescence to deviant stimuli ($SD = 33$%). The modal PPG result was 100%, as 15 subjects obtained full arousal to deviant sexual arousal with a range of 0% ($n= 1$) to 100% (median =38%).

A series of logistic regression analyses were used to test the hypotheses. First, logistic regressions allow examination of the hypotheses that the Static-99 score would predict both general and sexual offense reconviction. Regression results indicated the overall model fit of the predictor, Static-99 total score, was not significantly predictive of future conviction for any offense ($-2 \text{ Log Likelihood} = 111.72, \chi^2(1) = .241, p = .617$). The model correctly classified 85% of cases. When examining sexual reconviction, regression results indicated the overall model of fit using the Static-99 total score as a predictor was also not indicative of sexual reoffense ($-2 \text{ Log Likelihood} = 84.322, \chi^2(1) =$}
.410, \( p = .522 \)). This model correctly classified 90% of cases. As shown in Tables 1 and 2, the Static-99 did not predict new convictions for either any offense or sexual offenses at a statistically significant level.

Table 1

**Logistic Regression Coefficients – Conviction for Any New Offense, Static-99 Total Score**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Wald</th>
<th>df</th>
<th>p</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99</td>
<td>.07</td>
<td>.24</td>
<td>1</td>
<td>.617</td>
<td>1.07</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.89</td>
<td>18.90</td>
<td>1</td>
<td>.000</td>
<td>.15</td>
</tr>
</tbody>
</table>

Table 2

**Logistic Regression Coefficients – Conviction for a New Sexual Offense, Static-99 Total Score**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Wald</th>
<th>df</th>
<th>p</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99</td>
<td>.11</td>
<td>.42</td>
<td>1</td>
<td>.518</td>
<td>1.11</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.48</td>
<td>21.63</td>
<td>1</td>
<td>.000</td>
<td>.08</td>
</tr>
</tbody>
</table>

A second set of logistic regressions were calculated to determine the degree to which deviant sexual arousal predicted general and sexual offense convictions.

Regression results indicated the overall model fit of the predator, highest response on the PPG, was not predictive of future conviction of a criminal offense (-2 Log Likelihood = 111.96, \( \chi^2(1) = .001, p = .978 \)). The model correctly classified 85% of cases. Due to the low base rate of reoffense, the classification rate indicates that the model effectively
classified subjects that did not reoffend, however it was not accurate in predicting subjects that did reoffend. When examining sexual reconviction, regression results indicated that the overall model of fit using a dimensional model of the PPG as a predictor also was not predictive (-2 Log Likelihood = 84.53, $\chi^2(1) = .20, p = .656$). This model correctly classified 90% of cases and was again affected by the low base rate of reoffending in the sample. Coefficients for the model using PPG to predict any new conviction and conviction for a new sexual offense, respectively, are presented in Tables 3 and 4.

Table 3

*Logistic Regression Coefficients – Conviction for Any New Offense, PPG Results*

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Wald</th>
<th>df</th>
<th>p</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPG results</td>
<td>.00</td>
<td>.00</td>
<td>1</td>
<td>.978</td>
<td>1.00</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.72</td>
<td>16.25</td>
<td>1</td>
<td>.000</td>
<td>.18</td>
</tr>
</tbody>
</table>

Table 4

*Logistic Regression Coefficients – Conviction for a New Sexual Offense, PPG Results*

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Wald</th>
<th>df</th>
<th>p</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99</td>
<td>.00</td>
<td>.20</td>
<td>1</td>
<td>.655</td>
<td>1.00</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.40</td>
<td>20.26</td>
<td>1</td>
<td>.000</td>
<td>.09</td>
</tr>
</tbody>
</table>

To examine the relationship between the Static-99 total score and the PPG percentages, a bivariate correlation was conducted. Results showed a weak but
A statistically significant correlation between the Static-99 and PPG results in the expected direction ($r = .17, p = .030$). Parallel to the stated hypothesis, as subjects’ Static-99 score increased, the percent of deviant arousal also increased.

Stepwise multiple regressions were then conducted to determine whether deviant sexual arousal as measured by a dimensional model of the PPG added incremental validity to the Static-99 when predicting general and sexual offense convictions. Regression results indicate that the overall model including the variables of Static-99 and PPG results did not accurately predict general recidivism at a statistically significant level. A summary of regression models indicates that neither variable entered into the final regression equation and were not statistically significant contributors to predicting general recidivism ($R^2 = .04, R^2_{adj} = .00, F[1,128,] = .00, p = .956$). Similarly, Tables 6 and 7 show that the Static-99 and PPG results did not contribute to predicting conviction for a new sexual offense at a statistically significant level ($R^2 = .06, R^2_{adj} = .00, F[1,128,] = .12, p = .733$). These regression models indicate that the PPG does not add incremental validity to the Static-99 when predicting conviction for a new sexual or general offense.

Table 5

Model Summary – Conviction for Any New Offense

<table>
<thead>
<tr>
<th>Step</th>
<th>$R$</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>$\Delta R^2$</th>
<th>$F_{chg}$</th>
<th>$p$</th>
<th>$df_1$</th>
<th>$df_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Static-99</td>
<td>.04</td>
<td>.00</td>
<td>-.01</td>
<td>.00</td>
<td>.24</td>
<td>.624</td>
<td>1</td>
<td>129</td>
</tr>
<tr>
<td>2. PPG results</td>
<td>.04</td>
<td>.00</td>
<td>-.01</td>
<td>.00</td>
<td>.00</td>
<td>.956</td>
<td>1</td>
<td>128</td>
</tr>
</tbody>
</table>
Table 6

Coefficients for Final Model – Conviction for Any New Offense

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>β</th>
<th>t</th>
<th>p</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99 total score</td>
<td>.01</td>
<td>.04</td>
<td>.49</td>
<td>.624</td>
<td></td>
</tr>
<tr>
<td>PPG results</td>
<td>.00</td>
<td>-.01</td>
<td>-.06</td>
<td>.956</td>
<td>-.01</td>
</tr>
</tbody>
</table>

Table 7

Model Summary – Conviction for a New Sexual Offense

<table>
<thead>
<tr>
<th>Step</th>
<th>R</th>
<th>R²</th>
<th>R² adj</th>
<th>ΔR²</th>
<th>F chg</th>
<th>p</th>
<th>df₁</th>
<th>df₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Static-99</td>
<td>.06</td>
<td>.00</td>
<td>-.01</td>
<td>.00</td>
<td>.42</td>
<td>.52</td>
<td>1</td>
<td>129</td>
</tr>
<tr>
<td>2. PPG results</td>
<td>.06</td>
<td>.00</td>
<td>-.01</td>
<td>.00</td>
<td>.12</td>
<td>.77</td>
<td>1</td>
<td>128</td>
</tr>
</tbody>
</table>

Table 8

Coefficients for Final Model – Conviction for a New Sexual Offense

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>β</th>
<th>t</th>
<th>p</th>
<th>Partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99 total score</td>
<td>.01</td>
<td>.05</td>
<td>.58</td>
<td>.565</td>
<td></td>
</tr>
<tr>
<td>PPG results</td>
<td>.00</td>
<td>.03</td>
<td>.34</td>
<td>.733</td>
<td>.03</td>
</tr>
</tbody>
</table>

Independent sample t-tests were analyzed for a closer examination of those who recidivated versus those who did not reoffend. With regard to the Static-99 total score, there was no statistically significant difference between groups for sexual convictions.
(\(F[129] = .43, p = .513\)) or conviction for any new offense (\(F[129] = .13, p = .717\)).

Additionally, there was not a statistically significant difference between groups when comparing degree of deviant sexual arousal for those who recidivated sexually (\(F[129] = 1.01, p = .317\)) or with any conviction (\(F[129] = .00, p = .962\)). Between group comparisons are summarized in Tables 9 and 10.

### Table 9

*Independent Samples t-Test – Conviction for a New Sexual Offense*

<table>
<thead>
<tr>
<th></th>
<th>Mean Scores</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reoffended</td>
<td>No reoffense</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>((N=13))</td>
<td>((N=118))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static-99 total</td>
<td>2.77 (SD= 1.92)</td>
<td>2.44 (SD= 1.73)</td>
<td>.43</td>
<td>.64</td>
</tr>
<tr>
<td>PPG results</td>
<td>51.23 (SD= 36.76)</td>
<td>46.94 (SD= 32.61)</td>
<td>1.01</td>
<td>.45</td>
</tr>
</tbody>
</table>

### Table 10

*Independent Samples t-Test – Conviction for Any New Offense*

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reoffended</td>
<td>No reoffense</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>((N=20))</td>
<td>((N=111))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static-99 total</td>
<td>2.65 (SD= 1.66)</td>
<td>2.44 (SD= 1.76)</td>
<td>.13</td>
<td>.03</td>
</tr>
<tr>
<td>PPG results</td>
<td>47.55 (SD= 33.21)</td>
<td>47.33 (SD= 33.01)</td>
<td>.00</td>
<td>.49</td>
</tr>
</tbody>
</table>
To take into account the relatively low base rate of new charges, a Receiver Operating Characteristic (ROC) was analyzed to obtain the Area Under the Curve (AUC). As explained previously, AUC statistics are unaffected by base rates and a result of .50 equals chance, with greater values indicating predictive quality beyond chance. Deviant sexual arousal as measured by the PPG performed poorly as an independent predictor of conviction for a new sexual offense \((AUC = .52, \text{Asymptotic significance} = .79, CI = .35-.69)\), with results indicating a level of prediction no better than chance. The Static-99 also performed as poorly with this community sample, showing an AUC coefficient of .55 \((\text{Asymptotic significance} = .53, CI = .38-.73)\) when predicting conviction for a new sexual offense. When predicting general recidivism, PPG results also performed at chance \((AUC = .51, \text{Asymptotic significance} = .94, CI = .37-.64)\), as did the Static-99 \((AUC = .54, \text{Asymptotic significance} = .53, CI = .41-.68)\).
DISCUSSION

Overview of Findings

In this attempt to extend the utility of the Static-99 to community settings, results showed that this actuarial tool performed poorly with released sexual offenders, predicting recidivism at a level barely above chance. Results showed that the Static-99 was not predictive of either sexual or general recidivism beyond chance and mean risk scores were not statistically different between subjects who reoffended and subjects who did not recidivate. These findings were divergent from results purported in other studies where the Static-99 predicted sexual and violent recidivism well above chance (Hanson & Thornton, 2000).

Contrary to the stated hypotheses, percent of deviant sexual arousal as measured by PPG was not significantly predictive of sexual or conviction for any new offense. Furthermore, a dimensional model of the PPG results did not add incremental validity to the Static-99 when predicting sexual or general recidivism as was expected given previous use of deviant sexual arousal as a risk assessment item (Quinsey et al., 2007). Although encouraging that an elevated score on the Static-99 was also indicative of an increased percent of arousal to deviant sexual stimuli, these results did not translate into predicting recidivism. The number of subjects in the current study is similar to other studies examining the utility of the PPG (Howes, 1995; Launay, 1994); however it is difficult to compare the sample characteristics of PPG results across studies due to the
lack of reported information (e.g., mean level of arousal) and variability across studies in administration and interpretation techniques of PPG data.

Although many risk assessment measures can perform well in highly controlled validation studies, those results may not translate to other groups, especially if the group is dissimilar in some way. A similar explanation may contribute to the current results in regard to the Static-99, as few validation studies have been completed with participants in the United States to date, and even fewer have been performed with a community sample. The current community sample had similar base rates for sexual recidivism as has been recorded in previous studies (11.9%; Hanson & Mouton-Bourgon, 2009), but results of the current study indicate a substantially lower base rate for conviction for any new offense at 15% of the total sample versus 30 to 60% in the validation study (Hanson & Thornton, 2000). The mean Static 99 scores reported in several larger studies were substantially higher than the current sample ($M = 2.47$); earlier researchers reported means of 2.94 (Boccaccini et al., 2009) and 3.15 (Hanson & Thornton, 2000), indicating that subjects in the current study were at lower risk to reoffend. Unlike subjects in the original validation studies of the Static-99, many of the current subjects received long-term probation as a sentence and did not spend time in confinement. Some subjects in the current study had been investigated for, or charged with, sexual offenses, but after the sexual offender evaluation were ultimately not adjudicated specifically as a sexual offender (e.g., they entered a plea to lesser charges or charges were dropped after the investigation). These differences are important to note because there may be a distinction in recidivism risk between groups of offenders who have spent time in prison or civil commitment and those that remained in the community.
In regard to the PPG, percent of deviant sexual arousal did not differentiate between offenders who recidivated and those who did not reoffend. These nonsignificant results contribute to the variable results reflected in the literature that cast doubt on the utility of deviant sexual arousal as measured by the PPG to reliably contribute to risk prediction. Both the Static-99 and PPG results are used to make important decisions about restricting personal liberties to released offenders. Additionally, scores on these measures can affect an offender’s progression in treatment programs, often a condition of provisional release. Authorities and treatment providers basing decisions on the Static-99 or PPG data in a community setting may be led to erroneous or misleading conclusion about a sexual offender’s risk of reoffense.

Although the clinical and treatment utility of the PPG remains for future researchers to address, the current research indicates that increased deviant sexual arousal as measured by percentage of tumescence was not predictive of recidivism. This finding is contrary to the small amount of past research findings advertising the PPG as the most reliable and objective method of measuring the physiological arousal to deviant sexual material. Deviant sexual arousal was added to the SORAG in an effort to increase predictive accuracy for determining recidivism risk specifically for sexual offenders (Quinsey et al., 2006). However, the current study highlights that determining the most accurate and reliable method to measure the construct of deviant sexual arousal is still a matter of controversy.

Strengths of the Current Research

There were several strengths in the current research project. First, this unique community sample allowed for the exploration of the complex task of risk prediction
with a sample of individuals living in the community. As reflected in results that differed from previous research, community management and risk prediction with sexual offenders may be more complex than reflected in original validation samples of incarcerated or committed individuals. Comparing offenders on community release to validation populations that are confined or geographically separated may be inappropriate, which is important information for clinicians who need to use the most appropriate and accurate tools possible. The follow-up period utilized by this study was on par with other research studies. This study also mirrored the recommendations put forth by the authors of the Static-99 and utilized logistic regression as the primary means for assessing the effectiveness of the risk assessment measures (Helmus et al., 2009).

**Limitations of the Current Research**

Though procedures were put in place to control the input of data, including a code book and checks of random files, errors are to be expected because of inherent error associated with data collection involving multiple research assistants and many points of data. In a community setting, multitudes of confounding variables may have influenced the results of the present study. Unlike a setting in which participants are confined in a prison or hospital, the offenders in the current study were free to make lifestyle choices about factors such as employment status, association with confederates, substance use, and housing situations. Previous studies indicated that these dynamic variables influence an offender’s reoffense status (Hanson, 2004), and they may have affected conviction rates in the current study as well. Additionally, data obtained from the Department of Corrections were conviction data, not original arrest or charge information. This is problematic in two ways. First, many offenses are pled down to lesser degrees and may
not show as sexual offenses in the database. Second, sexual offenses often go undetected for years, and many are never reported to law enforcement. As with any study on sexual offense recidivism, the reporting and conviction rates of individuals may have played a part in the results of the current study. Victims of sexual assault are often discouraged from reporting their experience due to a myriad of factors, including social stigma, relationship to the perpetrator, and fear of repercussions (Bureau of Justice Statistics, 2002). The reoffense rate of the current study was most likely affected by this societal tendency of underreporting of sexual offences and thus would have reflected a low, if not inaccurate, reoffense rate. Additionally, a lack of racial and ethnic diversity in the sample may skew the results of the current study, especially considering the diverse make up of criminal justice populations.

Future Directions

Although the current study employed a respectable number of subjects to produce statistically significant results ($N = 131$), future researchers should focus on expanding the number of offenders included in the data. An effort should be made to expand the number of sexual offender treatment programs represented in the data to combat any bias for individual treatment programs to accept specific types of offenders or sexual offenders at an assumed level of risk. When exploring the PPG, new research should focus on the appropriate uses for the data obtained in a phallometric evaluation. Attention in research should be focused on the predictive abilities of the PPG for risk assessment to either confirm the results of the current study, which showed the PPG to perform at a chance level when predicting recidivism, or dispute the current findings. Future researchers may also indicate more effective and reliable ways of organizing and
interpreting PPG data in regard to risk assessments. Research should continue to be conducted on the Static-99 in different geographic areas and with diverse ethnic and racial populations as well as individuals with varying levels of contact with the criminal justice system in order to create useful normative statistics from which to compare offenders.

In general, given the current nonsignificant results, more studies need to be conducted to shed light on the conflicting evidence of using the Static-99 and PPG to predict risk of recidivism with sexual offenders. Future researchers should attempt to clarify the most appropriate settings, populations, and referral questions in which to use this actuarial tool and objective measure of deviant sexual arousal in order to further efforts towards differentiating those offenders at risk of reoffending and predicting recidivism. Additionally, methods may need to be developed to capture a more accurate and sensitive measurement of sexual recidivism, such as self-report measures, that do not rely on official reconviction data. Future researchers should concentrate on expanding the follow-up period due to the propensity for some sexual offenders to recidivate without being reported to provide a more accurate view of recidivism risk over time.

Conclusion

Based on the results of the current study, neither the Static-99 nor the PPG lived up to past research results touting their predictive qualities in addressing recidivism with sexual offenders. Basic principles of psychometrics demand that psychological tests be standardized, reliable, valid, and predictive of future behavior in order to be appropriately used in decision making (Fernandez, 2009). Though the Static-99 appears to meet the first three of these criteria, the question remains after the current study of if the measure
can reliably predict future behavior of individuals not incarcerated for sexual deviancy. The PPG appears to continue to struggle with standardization contributing to questions of the measure’s reliability and validity, and the current study reflects questions about the utility of the PPG in reliably predicting recidivism in a community sample. There is little consensus on how to interpret PPG data. Evaluation and treatment professionals differ in classification of deviant arousal from 5% to 30% tumescence with only vague guidelines from practicing associations (ATSA, 2004; Fernandez, 2009; Howes, 1995). Although the utility of the PPG results in treatment purposes was beyond the purview of the current study, the current research results indicate that data from the PPG fell well short of standards to be used in risk assessment. More effort toward standardization is needed for more widespread use and acceptance of the PPG as a method of risk assessment in the community (Fernandez, 2009). These measures are purportedly among the top tools for predicting risk among sexual offender populations in the community and many important decisions are made based on the scores individuals obtain. An at chance level of predicative quality of a risk assessment measure is unacceptable and falls far below practice standards in risk assessment with sexual offenders. The current results are indicative of a need for continued research and refinement of the Static-99 and PPG results in order to produce more reliable measures of recidivism.
REFERENCES


Association for the Treatment of Sexual Abusers (2004). *Practice standards and guidelines for the evaluation, treatment and management of adult male sexual abusers*. Beaverton, OR.


