

University of Windsor

## Scholarship at UWindor

---

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

---

Fall 2021

# Predicting Recidivism versus Desistance in Adolescents with Sexual Offenses Using Tools Developed for Adults: The Static-99 and SAPROF

Meredith Awrey  
*University of Windsor*

Follow this and additional works at: <https://scholar.uwindsor.ca/etd>



Part of the [Criminology Commons](#), and the [Psychology Commons](#)

---

### Recommended Citation

Awrey, Meredith, "Predicting Recidivism versus Desistance in Adolescents with Sexual Offenses Using Tools Developed for Adults: The Static-99 and SAPROF" (2021). *Electronic Theses and Dissertations*. 8888.

<https://scholar.uwindsor.ca/etd/8888>

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email ([scholarship@uwindsor.ca](mailto:scholarship@uwindsor.ca)) or by telephone at 519-253-3000ext. 3208.

**Predicting Recidivism versus Desistance in Adolescents with Sexual Offenses Using Tools**

**Developed for Adults: The Static-99 and SAPROF**

By

**Meredith Awrey**

A Thesis  
Submitted to the Faculty of Graduate Studies  
through the Department of Psychology  
in Partial Fulfillment of the Requirements for  
the Degree of Master of Arts  
at the University of Windsor

Windsor, Ontario, Canada

2021

© 2021 Meredith Awrey

PREDICTING RECIDIVISM VERSUS DESISTANCE IN ADOLESCENTS WITH SEXUAL  
OFFENSES USING TOOLS DEVELOPED FOR ADULTS: THE STATIC-99 AND SAPROF

by  
Meredith Awrey

APPROVED BY:

---

S. McMurphy  
Department of Sociology, Antropology, and Criminology

---

L. Erdodi  
Department of Psychology

---

C. Langton, Advisor  
Department of Psychology

August 27<sup>th</sup>, 2021

## DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

## ABSTRACT

The present study investigated the predictive validity of the risk-based Static-99 and strengths-based SAPROF with a sample of adolescent males ( $N = 272$ ) who have sexually offended. Additionally, the incremental validity of the SAPROF over the Static-99 was tested. Participants received a specialized assessment at one of three sites Southern Ontario. Archived case files were retrospectively coded using modified versions of the Static-99 and SAPROF. The Static-99 (modified), SAPROF total score (modified), and SAPROF internal and motivational (modified) subscales were found to significantly predict serious recidivism and the modified SAPROF total score, internal subscale, and modified motivational subscale were found to significantly predict any recidivism. These findings have implications for our understanding of risk and protective factors among adolescents who have offended sexually.

## ACKNOWLEDGEMENTS

First, I would like to thank my research advisor, Dr. Calvin Langton, for his invaluable support and guidance. His mentorship and confidence in me have made me a better and more confident researcher. I would also like to express my appreciation to Dr. Langton for allowing me to use some of the data he collected as a part of his grant-funded project for this thesis.

I would like to thank Dr. Laszlo Erdodi and Dr. Dana Levin for their time, as well as their valuable insight and contributions to this project. Additionally, I would like to thank Dr. Suzanne McMurphy for her contributions, time, and for kindly stepping in to join my committee.

I would also like to acknowledge that the data used in this thesis were coded from three specialized services, each of which has benefitted from the expertise of Dr. James Worling. I would like to express my gratitude to Dr. Worling, who is collaborating with Dr. Langton on the project from which the data in this thesis is a part of. Additionally, I would like to thank the clients and their families for their hard work.

Next, I would like to thank my cohort. I could not have made it this far without the support and laughter they have supplied. I'm so grateful to be sharing this journey with each one of you.

A special thank you to my best friend, Madison MacKinnon, for always pushing me on to greater things.

Finally, I would like to thank my parents, Laura and William Awrey, for always believing in me. I would not be where I am today without their endless encouragement and support of my academic endeavours.

TABLE OF CONTENTS

DECLARATION OF ORIGINALITY.....iii

ABSTRACT.....iv

ACKNOWLEDGMENTS.....v

LIST OF TABLES.....ix

Introduction.....1

    Overview.....1

    Adolescent Offending.....3

        Risk and Protective Factors.....3

    Assessing Likelihood of Reoffending.....7

        Actuarial Scales.....7

        Structured Professional Judgement Scales.....10

        Efficacy of Actuarial vs Structured Professional Judgement Methods.....12

    Risk-Based Assessment Tools.....13

        Challenges to Risk-Based Assessment.....16

    Strengths-Based Tools.....19

        Desistence.....20

    Research Design in Forensic Clinical Assessment.....21

        Statistical Methods.....23

    Demographic Considerations in Risk Assessment.....25

    The Use of The Static-99 and SAPROF in Adolescent Samples.....27

    The Static-99.....28

        Overview and Development.....28

|                               |    |
|-------------------------------|----|
| Use with Adults.....          | 43 |
| Predictive Validity.....      | 43 |
| Interrater Reliability.....   | 47 |
| Use with Adolescents.....     | 48 |
| The SAPROF.....               | 40 |
| Overview and Development..... | 40 |
| Use with Adults.....          | 43 |
| Predictive Validity.....      | 43 |
| Interrater Reliability.....   | 47 |
| Use with Adolescents.....     | 48 |
| Present Study.....            | 51 |
| Hypotheses.....               | 52 |
| Methods.....                  | 55 |
| Participants.....             | 55 |
| Measures.....                 | 56 |
| Static-99 (modified).....     | 56 |
| SAPROF (modified).....        | 57 |
| Outcomes.....                 | 57 |
| Procedure.....                | 57 |
| Data Analysis.....            | 58 |
| Hypothesis 1.....             | 58 |
| Hypothesis 2.....             | 59 |
| Hypothesis 3.....             | 59 |

|                              |    |
|------------------------------|----|
| Hypothesis 4.....            | 60 |
| Results.....                 | 60 |
| Inter-rater Reliability..... | 60 |
| Predictive Validity.....     | 61 |
| Incremental Validity.....    | 64 |
| Model 1.....                 | 65 |
| Model 2.....                 | 65 |
| Model 3.....                 | 66 |
| Model 4.....                 | 66 |
| Model 5.....                 | 67 |
| Model 6.....                 | 67 |
| Discussion.....              | 71 |
| Limitations.....             | 78 |
| Future Directions.....       | 78 |
| Conclusion.....              | 80 |
| References.....              | 82 |
| Vita Auctoris.....           | 99 |

## LIST OF TABLES

|                                                                                                             |    |
|-------------------------------------------------------------------------------------------------------------|----|
| • <i>Table 1.</i> Static-99 Items.....                                                                      | 30 |
| • <i>Table 2.</i> Summary of Key Findings with Regards to the Use of the Static-99 with<br>Adolescents..... | 39 |
| • <i>Table 3.</i> SAPROF Items.....                                                                         | 42 |
| • <i>Table 4.</i> Summary of Key Findings with Regards to the Use of the SAPROF with<br>Adolescents.....    | 51 |
| • <i>Table 5.</i> Hypotheses, Justification, and Corresponding Data-Analytic Plan.....                      | 54 |
| • <i>Table 6.</i> Descriptive Statistics and Inter-rater Reliability.....                                   | 61 |
| • <i>Table 7.</i> Prediction of Serious Recidivism.....                                                     | 63 |
| • <i>Table 8.</i> Prediction of Any Recidivism.....                                                         | 64 |
| • <i>Table 9.</i> Hierarchical Models for Serious Recidivism.....                                           | 69 |
| • <i>Table 10.</i> Hierarchical Models for Any Recidivism.....                                              | 70 |

## Introduction

### Overview

Adolescent sexual offending is a prominent issue, with youth making up 17% of all individuals accused of a sexual offense in Canada (Allen & Superle, 2016). Sexual violence has notable costs for both victims and broader society. Victims experience negative mental (Austin & Short, 2020; Hanson, 1990; Vandemark & Mueller, 2008) and physical health consequences (Waigandt et al., 1990) after experiencing sexual violence. Further, society as a whole incurs costs, including monetary losses associated with sexual offending. For example, in Canada in 2009, costs associated with sexual offenses totaled \$4.8 billion, including justice system and medical costs, lost productivity, employer losses, and social services (Hoddenbagh et al., 2014). Given this, there is a clear need to attend to adolescent sexual offending and sexual recidivism.

There is a plethora of tools intended to assess the likelihood of reoffending, some of which are specifically intended for assessing likelihood of sexual reoffending (Boer et al., 1997; Grubin, 1998; Hanson & Bussiere, 1998; Hanson & Thornton, 2000; Quincey et al., 2006), and some which are specifically intended for use with youth (de Vries Robbé et al., 2015; Epperson et al., 2006; Prentky & Righthand, 2003; Worling & Curwen, 2001). Given that there are both ethical (American Psychological Association [APA], 2013, Section 1.02, 11.01; Canadian Psychological Association [CPA], 2017, Section II.1, II.2, II.18) and developmental concerns (Barbaree et al., 2009; Hanson, 2002; Hanson & Thornton, 2000) associated with using tools intended for a specific population with individuals outside of that demographic, it is necessary to ensure that risk assessment tools have been shown to have predictive validity in a given population before they are used.

The present study aimed to determine whether a modified version of the widely used and researched risk-based assessment tool intended for use with adults, the Static-99 (Hanson & Thornton, 2000), and a modified version of a newer, strengths-based assessment tool also intended for use with adults, the Structured Assessment of Protective Factors for violence risk (SAPROF; de Vogel et al., 2011), have predictive validity for two distinct recidivism outcomes (a new serious offense and any new offense during follow-up) in a sample of adolescents who have sexually offended. The modifications of each involve the omission of items not codable with adolescents. Further, given that strengths-based measures may be incrementally valid over risk-based measures (de Vries Robbé et al., 2011, 2013, 2015b) the present study also aimed to determine whether the modified SAPROF has incremental validity over the modified Static-99.

The data for this study was drawn from an existing dataset collected by Dr. Langton as part of his larger SSHRC-funded grant project. The sample for the study was an archival one comprised of more than 300 male adolescents referred over an 18-year period to specialized services for youth with sexual offenses in a major urban area in Southern Ontario. Within that larger dataset the tools had been coded (blind to criminal records follow-up information) from the archived case files for the sample. Official criminal records information was also available in the larger dataset with both sexual and any recidivism outcomes coded dichotomously for the follow-up period, an average of 5.9 years for the sample. ROC analyses and Cox regression survival analyses were used to determine if there are main effects and incremental validity over the modified Static-99 for the modified SAPROF protective factors with adolescents with sexual offenses. The findings from this study contribute to the literature on recidivism versus desistance among youth justice-involved adolescents, highlighting developmental considerations with implications for assessment and intervention work.

## **Adolescent Offending**

In 2019, approximately 77,000 Canadian youth between the ages of 12 and 17 were accused of a criminal offense by police. The crimes most frequently committed by youth include assault, shoplifting under 5000 dollars, and mischief. Approximately 32,000 youth were accused of a violent crime in 2019, and approximately 3,500 were accused of a sexual offense, including sexual violations against children (Moreau, 2019). In 2014, 17% of all individuals accused of sexual offenses were youth, and individuals age 14 to 16 were accused of sexual assault level one at a higher rate than any other age group. In cases where a youth was accused of a sexual offense, the victim was 12 years old or younger 41% of the time (Allen & Superle, 2016). It is clear that youth perpetrate sexual offenses at a notable rate, and thus it is essential for both policy and practice that we address sexual offending by adolescents. The present study aimed to address specifically the assessment of recidivism risk in adolescents who have sexually offended.

### ***Risk and Protective Factors***

Underpinning the development and evaluation of tools intended to assess likelihood of reoffending, there is a notable body of research investigating risk factors which predict reoffending. In adults who have sexually offended, a plethora of risk factors for sexual recidivism have been identified (Arnold & Davis, 2016; Craig et al., 2005; Hanson & Bussiere, 1998; Hanson & Harris, 2000; Thornton, 2013). Static risk factors, or those which do not change, include prior sexual offenses, early onset of sexual offending, victims who are strangers, male victims, multiple victims, psychopathy, young age, being single and never married, experience of sexual abuse in childhood, and prior general offending (Arnold & Davis, 2016; Craig et al., 2005). Dynamic risk factors, which can change or be targeted by treatment, include difficulties with impulse control and self-regulation, substance abuse, poor problem-solving skills, social

skills deficits, poor motivation for treatment or failure to complete treatment, attitudes which are tolerant of sexual offending, and social isolation (Arnold & Davis, 2016; Craig et al., 2005; Hanson & Harris, 2000; Thornton, 2013). Some risk factors appear to be dynamic but may or may not be targeted and altered with treatment, and thus may actually be static in some individuals. These are largely related to sexual deviance and include sexual interest or preference for children or rape, and presence of multiple paraphilias (Arnold & Davis, 2016; Craig et al., 2005; Thornton, 2013).

Risk factors associated with a greater likelihood of reoffending among adolescents who have sexually offended certainly overlap with risk factors for adults. Specifically, static factors such as experiencing sexual abuse, multiple victims, prior sexual offenses, victims who are strangers, and male victims, and dynamic factors such as poor self-control and impulsivity, social isolation, poor motivation for change, uncompleted treatment, and attitudes which support sexual offending are associated with increased risk for reoffending in adolescence as well as adults (Arnold & Davis, 2016; Craig et al., 2005; Efta-Breitbach & Freeman, 2004; Hanson & Harris, 2000; Långström, 2002; Thornton, 2013; Worling & Långström, 2003, 2006). As with adults, factors which may be static or dynamic, such as sexual interest in prepubescent children or sexual violence, are also associated with a greater likelihood of reoffending in adolescents (Worling & Långström, 2003, 2006).

In addition to those which are also risk factors for adults, some risk factors specific to adolescents who have sexually offended have been identified. These include static factors such as parental divorce or living in a single parent household, and having siblings who have been sexually abused, as well as dynamic factors such as truancy, highly stressful family environments, and negative parent-child relationships (Worling & Långström, 2003, 2006).

Similarly, some risk factors which are relevant to adults who have sexually offended would not be developmentally suitable for consideration with adolescents, such as those relating to marital status (Worling & Långström, 2006). Given the presence of unique risk factors for adolescents as well as some risk factors which are only suitable for consideration with adults, there is a clear need to consider the age of an individual who has offended when assessing risk for reoffending. Additionally, since, as previously noted, scales intended to assess recidivism risk are developed based on risk factors known to predict recidivism, there is also a need to consider target age when developing risk assessment tools and ascertaining their appropriateness for use with a given individual. Further, assessment of strengths in addition to risks is necessary, as tools comprised of predictive factors may be statistically more accurate at predicting outcomes than tools comprised of risk factors and/or have incremental validity over risk risk-based tools (de Vries Robbé et al., 2011, 2013, 2015b).

Consideration of an individual's strengths or protective factors is essential when assessing the likelihood that they will reoffend. Despite this, there is a notably smaller body of literature investigating such factors, as compared to the research focused on risk. Strengths may be conceptualized as factors which reduce risk, factors which are protective specifically in the presence of risk factors, or as the opposite of risk factors (Thornton, 2013). It should be noted that although strengths and protective factors are frequently used interchangeably, there is a distinction to be made between the two. Specifically, strengths can be understood as positive external and internal resources that an individual has access to (Wanamaker et al., 2018). Protective factors, which can be conceptualized as a type of strength, refer to factors which buffer negative outcomes in high-risk individuals, and thus have a less pronounced effect in lower risk groups (Wanamaker et al., 2018). Thus, strengths are variables that are desirable and

are associated with more positive outcomes, and protective factors are variables that provide a buffer against negative outcomes in high-risk individuals (Wanamaker et al., 2018). The presence of protective factors is associated with a decreased likelihood of reoffending, or in other words, a greater likelihood of desistance from recidivism (Lodewijks et al., 2010). Among adults who have sexually offended, factors which are protective against recidivism include professional support, well-functioning individuals in one's social network, having an emotionally intimate confidante, structured prosocial group activities, such as employment, education, and leisure, problem solving skills, a sense of personal agency, and a belief that desistance is both possible and worthwhile (Thornton, 2013). Protective factors for adolescents who have sexually offended include positive peer relationships, above average cognitive abilities, coping skills, high socio-economic status, and a positive, affectionate, and supportive family environment (Efta-Breitbach & Freeman, 2004).

As noted when discussing risk factors, some protective factors are unique to adolescents, such as those related to the family environment, and thus indicate a need for consideration of an individual's age when assessing their likelihood of reoffending as well as when determining the appropriateness of a given strengths-based assessment tool for use with an individual who has offended. Specifically, given that both risk and protective factors demonstrably differ to some extent across age groups, it cannot be assumed that an assessment tool, based on known risk or protective factors for one age group, will accurately assess recidivism risk in another age group. An understanding of both risk and protective factors, including how they differ across age groups, is essential since, as previously noted, both actuarial and structured professional judgement scales are developed based on research regarding the validity of various factors as predictors of reoffending (Harris & Hanson, 2010; Rettenberger & Hucker, 2011).

## **Assessing Likelihood of Reoffending**

Over the past 30 years, there has been a proliferation of risk assessment tools intended to assess the likelihood of criminal reoffending. There are different categories of risk assessment tools or approaches to risk assessment typically described in the literature: unstructured professional judgement (which involves the clinician making a judgement about an individual's risk level based on their own knowledge and intuition), static actuarial scales, and dynamic actuarial scales (Douglas, 2019; Harris & Hanson, 2010; Rettenberger & Hucker, 2011). Finally, alongside the development of actuarial scales, particularly in more recent years, structured professional judgement tools have been developed (Douglas, 2019; Guyton & Jackson, 2008; Rettenberger & Hucker, 2011). An understanding of the distinction between these forms of assessment tools is essential, as the scales which are the focus of the present study are different types of assessment scales. Specifically, the Static-99 (Hanson & Thornton, 2000), which in its current iteration is the Static-99R (Phenix et al., 2016), is an actuarial scale and the SAPROF (de Vogel et al., 2011) is a structured professional judgement tool.

### ***Actuarial Scales***

Early efforts to assess risk of recidivism (i.e., the likelihood that an individual was dangerous, to use an antiquated label) involved unstructured professional judgement, which relies solely on the knowledge and experience of the clinician conducting the assessment (Douglas, 2019; Harris & Hanson, 2010; Guyton & Jackson, 2008; Rettenberger & Hucker, 2011). This method is generally considered to be the least useful form of risk assessment, as it is not comprehensive or transparent regarding the factors which lead to the determination of an individual's level of risk (Guyton & Jackson, 2008; Rettenberger & Hucker, 2011). Further, unstructured professional judgement is prone to error from bias, is not well validated empirically,

and ultimately tends to be inaccurate (Bengtson & Långström, 2007; Mori et al., 2017). A second category of forensic risk assessment methods, actuarial assessment tools, were developed with the aim of addressing these limitations (Douglas, 2019; Guyton & Jackson, 2008; Rettenberger, & Hucker, 2011).

The second approach to the assessment of risk of recidivism involves the use of actuarial scales assessing static, or unchanging risk factors (Douglas, 2019; Guyton & Jackson, 2008; Harris & Hanson, 2010; Rettenberger & Hucker, 2011; Rich, 2009). Actuarial measures are developed statistically by combining a number of risk factors empirically shown to predict recidivism into a highly structured scale (Douglas, 2019; Guyton & Jackson, 2008; Harris & Hanson, 2010; Kroner, 2019; Rettenberger & Hucker, 2011; Rich, 2009). It should be noted that, given that actuarial scales consist of risk factors empirically shown to predict recidivism, and given that risk factors differ to some extent across the age range, it is important to attend to the age group of the sample that a given scale is based on. Without sufficient empirical support indicating the scale is an accurate predictor of recidivism in another age group, this cannot be assumed.

An early example of a static, actuarial scale that has been subject to extensive replication work and achieved widespread use in applied practice was the Violence Risk Appraisal Guide (VRAG; Harris et al., 1993). The Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR; Hanson, 1997) which includes four static risk factors, and the Structured Anchored Clinical Judgement – Minimum (SACJ-Min; Grubin, 1998) were developed shortly after (Harris & Hanson, 2010). The Static-99, one of the most widely used actuarial risk assessment tools (Archer et al., 2006), was created via the combination of the RRASOR and SACJ-Min, and consists of 10 static risk factors (Hanson & Thornton, 2000). The Static-99 was revised into the

Static-99R with the alteration of the age item, which had previously been coded dichotomously with age below 25 years indicating a risk and age above 25 years as a non-risk (Hanson & Thornton, 2000; Helmus et al., 2012). In the Static-99R, the age item was broken down into four categories: 18-34.9, 35-39.9, 40-59.9, and above 60 (Helmus et al., 2012; Phenix et al., 2016).

Static risk-based measures have been criticized for lacking clinical utility, as static risks are unchangeable and cannot be targeted in treatment (Douglas, 2019; Harris & Hanson, 2010; Kroner, 2019). This resulted in the development of a number of actuarial measures consisting of dynamic risk factors, or the third category of risk assessment tools (Harris & Hanson, 2010; Rettenberger & Hucker, 2011). These scales consist of risk factors that are empirically shown to predict reoffending in the same manner as those of the static risk assessment scales; however, the factors included in these scales are dynamic and therefore can be targeted in treatment (Harrison & Hanson, 2010). Examples of these scales include the STABLE-2007 and ACUTE-2007 (Hanson et al., 2007). These scales are intended to predict changes in recidivism risk among individuals who have sexually offended (Harris & Hanson, 2010). As its name suggests, the STABLE-2007 is intended to assess stable, or static risk factors across five categories: significant social influences, intimacy deficits, sexual self-regulation, general self-regulation, and cooperation with supervision (Hanson et al., 2007; Hanson & Harris, 2010). The ACUTE-2007 consists of eight items which are intended to assess current behaviour, such as hostility, sexual preoccupation, substance abuse, and rejection of supervision (Hanson et al., 2007; Harris & Hanson, 2010).

Actuarial scales have come under criticism. Perhaps the most noteworthy is that many of the most widely used actuarial risk assessment tools do not include clinical information which is likely highly relevant to an individual's risk of reoffending, such as psychological states,

interests, and drives (Douglas, 2019; Craig et al., 2005; Kroner, 2019). Given this, some have argued for the use of structured professional judgement scales, as they are comprised of empirically supported risk factors but also include an element of clinical judgement which may capture factors missed by actuarial scales (Guyton & Jackson, 2008; Rettenberger & Hucker, 2011).

### ***Structured Professional Judgement Scales***

Structured professional judgement scales are intended to capture a middle ground between unstructured professional judgement and actuarial tools, in which clinically relevant and idiosyncratic factors are considered, but empirically validated risk factors are also included (Douglas, 2019; Guyton & Jackson, 2008; Rettenberger & Hucker, 2011). These measures include explicit risk factors which are operationalized in a manual and scored in a standard manner (typically, 0 indicates the factor is not present for the individual; 1 indicates it is possibly or partially present; 2 indicates it is definitely present). In contrast, the scoring or weighting of items comprising actuarial scales is determined empirically. Once the full set of items comprising the structured professional judgement scale has been scored, the assessor then determines which items to weigh more heavily for the individual being assessed in order to conclude with a summary risk judgement (typically, that the individual is at low risk, moderate risk, or high risk of reoffending). Conversely, possible summed score ranges on an actuarial scale are associated with the percentage of those with scores in the same range who reoffended in the development or cross-validation samples used to develop or test the actual scale's predictive validity. Structured professional judgement methods tend to be based on three principles: prevention of future offending, rather than exclusive prediction of it, structure to organize the

work of clinicians, and flexibility in order to allow for individualized and contextualized risk assessment (Douglas, 2019; Guyton & Jackson, 2008; Rettenberger & Hucker, 2011).

Scoring of these measures typically involves six steps (Douglas, 2019; Logan, 2016). First, information is gathered from sources such as the case file and the individual themselves, if possible or necessary. Second, clinicians determine whether each of the empirically derived risk factors are present for that individual. Third, using clinical judgement, clinicians determine the extent to which each present risk factor is relevant for that specific individual's likelihood of reoffending. Fourth, information about the relevant risk factors is supplemented with clinical judgements about other potentially relevant factors and combined to form a case formulation about the individual's likelihood of reoffence. In the fifth step, risk management strategies are identified. Finally, in the sixth step, the clinician makes a judgement about the individual's level of risk based both on the empirically derived risk factors and any other clinical information and judgements they have gathered (Douglas, 2019; Logan, 2016). Given that the items making up these scales are empirically derived, consideration of the age group of the sample the items were developed with continues to be essential due to both risk and protective factors varying for different age groups. Thus, these scales can only be considered appropriate for use with the age group they were developed for in the absence of empirical evidence suggesting otherwise.

An early example of a structured professional judgement tool was the Sexual Violence Risk – 20 (SVR-20; Boer et al., 1997). It includes 20 sexual violence risk factors, including those relating to psychosocial adjustment, history of sexual offenses, and plans for the future. The HCR-20 is a widely used (Archer et al., 2006) structured professional judgement risk assessment tool, consisting of 20 items across three subscales: historical, clinical, and risk management (Webster et al., 1997). It is intended to assess risk for violent reoffending and be used as a guide

for structured clinical risk assessment (Webster et al., 1997). More recently, the Structured Assessment for Protective Factors for violence risk (SAPROF; de Vogel et al., 2011) was introduced. This tool includes 17 factors which protect against future violent offending, across three subscales: internal, external, and motivational (de Vogel et al., 2011).

The SAPROF is one of a small number of tools assessing likelihood of reoffending (really, the absence of reoffending or desistance from offending) by focusing on strengths rather than risks. It is intended for use in conjunction with a risk-based measure, such as the aforementioned HCR-20 (de Vogel et al., 2011). There is a notable trend in the literature on recidivism to focus on risks, both in terms of forensic assessment tools and factors predicting recidivism in general. This may be related to a tendency to view strengths or protective factors as mere opposites of risk factors (Heffernan & Ward, 2017) and thus as not providing additional meaningful information. Regardless, further research on strengths and strengths-based tools is needed to address this gap.

### ***Efficacy of Actuarial versus Structured Professional Judgement Methods***

There have been mixed findings regarding whether actuarial or structured professional judgement tools more accurately predict recidivism in individuals who have offended (Hanson & Morton-Bourgon, 2009; Singh et al., 2011). In a meta-analysis conducted by Singh and colleagues (2011), 68 studies of violence risk assessment tools were analyzed. The majority of studies included had samples consisting of adult males who had violently offended. No evidence was found indicating that either actuarial or structured professional judgement more accurately predicts violent recidivism risk (Singh et al., 2011). Although these findings may indicate that there is no difference in the accuracy of actuarial versus structured professional judgement risk assessments for violent recidivism, more evidence is required in order to conclude that these

findings extend to sexual recidivism, given that only studies investigating violent reoffending were included.

Hanson and Morton-Bourgon (2009) conducted a meta-analysis on 118 studies of recidivism risk assessments tools with individuals who sexually offended. Most studies in the analysis had adult male samples and used an average follow up period of 70 months. They found that structured professional judgement tools were intermediately accurate at predicting recidivism. More specifically, they found that actuarial assessment was the most accurate, followed by structured professional judgement, and unstructured professional judgement was the least accurate. It is important to note that this analysis included only six studies with structured professional judgement tools (Hanson & Morton-Bourgon, 2009), and thus its results may not accurately represent the differences in efficacy of these methods. Given that the mixed findings regarding the efficacy of actuarial versus structured professional judgements tools in assessing recidivism risk, further research comparing the two types of measures is needed in order to determine whether one is more accurate than the other. In addition to being classified as structured professional judgement or actuarial, forensic assessment scales may be risk or strengths based. A review of scales in each of these categories is provided below.

### **Risk-Based Assessment Tools**

As previously mentioned, there are a notable number of risk-based tools for the assessment of likelihood of reoffence. These tools are often intended for use with a specific population, predicting a specific type of reoffence, or in some cases, both. As previously noted, attention to the specific population a tool is intended for is essential, and empirical validation is necessary to determine if a tool is an accurate predictor of recidivism when used with populations outside of a tool's intended demographic.

Tools intended for specific populations, but predicting general or all types of recidivism, include the Level of Service/Case Management Index (LS/CMI; Andrews et al., 2000), and the Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge et al., 2005). The LS/CMI is a structured professional judgement instrument, intended for use with adults and adolescents above the age of 16 (Andrews et al., 2000). Treatment needs are also assessed, and a case management plan is developed based on the findings of this tool (Andrews et al., 2000). The YLS/CMI is intended specifically for use with youth who have offended. As with the LS/CMI, the client's needed level of service and a case management plan are identified (Hodge et al., 2002).

A large number of risk-based tools have an intended population with regards to the type of offense an individual has committed. Typically, these tools also are intended to assess for a specific type of reoffending, such as violent reoffending among individuals who have sexually offended. The Sex Offender Risk Appraisal Guide (SORAG; Quincey et al., 2006) is an actuarial tool intended for assessing risk for violent (including sexually violent) recidivism in adults who have sexually offended (Quincey et al., 2006). The Juvenile Sex Offender Assessment Protocol (J-SOAP-II; Prentky & Righthand, 2003) and Juvenile Sexual Offense Recidivism Risk Assessment Tool -II (JSORRAT-II; Epperson et al., 2006) are both actuarial measures intended for assessing risk for sexual recidivism in adolescents who have sexually offended. They differ in that the JSORRAT-II exclusively assesses static risk factors, whereas the J-SOAP-II assesses both static and dynamic risk factors (Epperson et al., 2006; Prentky & Righthand, 2003). Similarly, the Estimate of Adolescent Sexual Offense Recidivism (ERASOR; Worling & Curwen, 2011) assesses sexual recidivism risk among adolescents who have sexually offended; however, the ERASOR is a structured professional judgement tool.

The Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR; Hanson, 1997), Structured Anchored Clinical Judgement (SACJ-Min; Grubin, 1998), and Static-99 (Hanson & Thornton, 2000) are all actuarial tools focusing on static risk factors intended for assessing sexual recidivism risk in adults who have sexually offended. The Static-99, as previously noted, was developed by combining the RRASOR and SACJ-Min (Hanson & Thornton, 2000). The Sexual Violence Risk – 20 (SVR-20; Boer et al., 1997) is structured professional judgement tool which, like the Static-99, RRASOR, and SACJ-Min, is intended for use with adults who have sexually offended in order to assess risk for sexual recidivism. Finally, the Risk for Sexual Violence Protocol (RSVP; Hart et al., 2003) is a structured professional judgement tool which was developed to replace the SVR-20 and is also intended to assess risk for sexual recidivism. The RSVP is intended to be a more robust assessment of recidivism risk, including planning to manage high-risk scenarios (Hart et al., 2003).

Finally, there are a few risk-based tools intended to assess the likelihood of a specific type of reoffence with any individual who has offended. The Violence Risk Appraisal Guide (VRAG) is an actuarial tool specifically intended for assessing violent reoffending in any adult who has offended (Harris et al., 1993). The Historical, Clinical, And Risk Management Scales (HCR-20; Webster et al., 1997) are also intended for assessing violent reoffending in any adult who has offended. This tool is a structured professional judgement scale and is widely used in recidivism research (Webster et al., 1997). The Structured Assessment of Violence Risk in Youth (SAVRY) is structured professional judgement tool developed for the assessment of violent reoffending risk in any youth who has offended (Borum et al., 2006). Evidently, there are a plethora of recidivism risk assessment tools available, intended for use with specific populations, in order to assess risk for a specific type of reoffending, or both. Although it is important to

assess an individual's risk for reoffending, as previously mentioned, assessment of one's strengths in addition to risks may be advantageous both clinically and in terms of predictive accuracy (Lodewijks et al., 2010; Thornton, 2013).

### ***Challenges to Risk-Based Assessment***

Although the forensic clinical assessment research heavily focuses on risk factors for recidivism over strengths or protective factors, and there are many more risk-based tools than strengths-based tools, assessment of an individual's strengths in addition to risks is advantageous in a number of ways. First, inclusion of strengths aligns with positive psychology, which has been shown to be an efficacious approach to intervention (Parks et al., 2015; Parks & Schueller, 2014). Positive psychology involves a focus on well-being and one's strengths in clinical practice, particularly in the development and implementation of interventions (Parks et al., 2015; Parks & Schueller, 2014). Positive case conceptualizations for clients who have offended aim to focus on opportunities and strengths rather than deficits (Passmore & Oades, 2015). Ultimately, strengths should be included in assessment, treatment planning, and intervention when assessing individuals who have offended, rather than using a deficits-based approach in which only risk-factors are considered (Passmore & Oades, 2015).

The Good Lives Model (GLM) constitutes a positive psychology approach to rehabilitation for individuals who have offended with notable implications for the assessment of recidivism risk (Ward & Fortune, 2013; Ward & Mann, 2004). The title "good" refers not to moral goodness, but rather to primary and secondary goods all individuals are motivated to seek. Primary goods refer to outcomes which are sought for their own sake and contribute to an individual's well-being (Ward & Fortune, 2013; Ward & Mann, 2004). These include knowledge, mastery, relationships with others, and healthy living. Secondary goods are those

which can be used to obtain primary goods, such as joining a sports team in order to obtain the primary goods of mastery and friendships (Ward & Mann, 2004).

GLM posits that all individuals are motivated to pursue primary goods to some extent, and that individuals who have offended have done this using ineffective and socially unacceptable approaches (Ward & Fortune, 2013; Ward & Mann, 2004). Risk assessment under the GLM model is strengths-based. It includes consideration of the individual's motivations, achievements and their prioritization of primary goods (Ward et al., 2007; Ward & Fortune, 2013). Further, it should include assessment of which primary and secondary goods the individual currently possesses, and which they aspire to obtain. Risk factors should also be included in assessment, and thus assessment following this model includes both strengths and risks in its determination of the likelihood an individual will reoffend (Ward et al., 2007; Ward & Fortune, 2013).

Beyond positive psychology, inclusion of strengths-based tools may improve accuracy of forensic risk assessments. Specifically, strengths, or protective factors, have been found to buffer the impact of risk factors on an individual's likelihood of reoffence (Jones et al., 2015, 2016; Lodewijks et al., 2010). For example, Lodewijks and colleagues (2010) found that, among a sample of adolescent males who have offended and are categorized as high risk for violent reoffending, the rate of violent reoffending was significantly higher among individuals without protective factors as compared to those with protective factors. This indicates that when individuals are at high risk for recidivism and also have a high number of protective factors, the protective factors may buffer the impact of the risk factors and lower overall risk (Lodewijks et al., 2010). Given this, the inclusion of a strengths-based tool is necessary for an accurate assessment of likelihood of reoffence. An individual may be deemed highly likely to reoffended

with the use of a risk-based tool; however, they may actually have a lower likelihood of reoffending, which may have been captured if strengths-based tools were used.

Although some have argued that the assessment of strengths is unnecessary because they are merely the opposite of risk factors (Heffernan & Ward, 2017), strengths may have incremental validity over risk factors. A number of studies have found that the SAPROF has incremental validity over risk-based tools in predicting sexual and general recidivism in adult samples (de Vries Robbé et al., 2011, 2013, 2015b). For example, in a sample of adult male patients in a forensic psychiatric hospital, evidence for the incremental validity of the SAPROF over the HCR-20 was found (de Vries Robbé et al., 2013). More specifically, when the SAPROF was added to a hierarchical logistic regression model in which the HCR-20 had been included in the first step, the model was significantly improved. This indicates that the SAPROF may have incremental validity over a risk-based measure, the HCR-20, for assessing the likelihood an individual will violently reoffend (de Vries Robbé et al., 2013).

This finding has not been consistently reported in the literature, with several studies failing to find evidence for the incremental validity of the SAPROF (de Vries Robbé et al., 2016, Zeng et al., 2015). This inconsistency underscores the need for further research in order to determine whether strengths-based tools in general, and the SAPROF in particular, have incremental validity over risk-based tools. Given that the inclusion of strengths-based tools in forensic risk assessment is supported by positive psychology (Parks et al., 2015; Parks & Schueller, 2014; Passmore & Oades, 2015), and may improve the accuracy of the assessment (Jones et al., 2015, 2016; Lodewijks et al., 2010), and that strength-based tools may have incremental validity over risk-based tools (de Vries Robbé et al., 2011, 2013, 2015b), additional brief comments about such tools is warranted.

## **Strengths-Based Tools**

Although there are many risk-based tools available to assess likelihood of reoffending, there is a dearth of strengths-based tools, or those based on protective factors. Beyond a lack of tools including items assessing strengths or protective factors, there are even fewer tools available that exclusively focus on strengths, as a notable proportion of tools which assess strengths also assess risk factors. Despite the limited number of these tools available, these tools can also be categorized by whether they are intended for use with a specific population, to predict desistance from a particular type of reoffence, or both. Four are mentioned briefly here. The Youth Assessment and Screening Instrument (YASI; Orbis Partners, 2000) is a population-specific actuarial measure. It includes both strengths and risk factors and is intended for use with youth who have committed any offense in order to predict general reoffending. The Structured Assessment of Violence Risk in Youth (SAVRY; Borum et al., 2006) is an offense specific structured professional judgement tool, intended for use with any adolescent who has offended in order to assess likelihood of violent reoffence. As with the YASI, the SAVRY includes both risk and protective factors.

The Structured Assessment of Protective Factors for violence risk (SAPROF; de Vogel et al., 2011) and SAPROF – Youth Version (SAPROF-YV; de Vries Robbé et al., 2015) are structured professional judgment tools that exclusively assess strengths in order to ascertain likelihood of desistance from violent offending. The SAPROF is intended for use with adults who have offended and, as its name suggests, the SAPROF-YV is intended for use with youth who have offended. The SAPROF-YV includes four domains: resilience (social competence, coping, self-control, and preservation), motivational (future orientation, motivation for treatment, attitudes towards agreements and conditions, medication, school/work, and leisure activities),

relational (parents/guardians, peers, and other supportive relationships), and external (pedagogical climate, professional care, and court order) (de Vries Robbé et al., 2015). There is some evidence indicating that the SAPROF-YV is a valid predictor of violent and general recidivism in youth who have offended (Chu et al., 2020; Kleeven et al., 2020). Unlike risk-based tools, which assess likelihood of reoffending, strengths-based tools assess likelihood of desistance. This distinction is important, as it informs interpretation of findings related to the predictive validity of strengths-based tools.

### *Desistence*

Risk-based tools are intended to inform the prediction that an individual will reoffend. In contrast, tools that are comprised exclusively of strengths are intended to predict desistance from offending. In other words, these tools are used to assess the likelihood that a given individual will not reoffend. Desistence is generally conceptualized in one of two ways: a dichotomous variable in which an individual is found to have not reoffended at any point over a specific period of time, or a process over time (Shapland & Bottoms, 2017). The process over time approach to defining desistance is commonly found in the field of criminology (Bushway & Paternoster, 2013; Paternoster & Bachman, 2017; Shapland & Bottoms, 2017). With this framework, the process through which the frequency of an individual's offending is decreased until it eventually stops reflects desistance (Shapland & Bottoms, 2017).

It is the dichotomous conceptualization of desistance that was adopted for the present study and that is used in research examining strengths or protective factors in tools and the predictive validity of strengths-based measures. The studies in which the development and validation of strengths-based measures, such as the SAPROF, have been reported operationalized desistance one of two possible binary outcomes (absence of reoffending) (de Vogel et al., 2011;

Yoon et al., 2018). This is consistent with studies evaluating risk assessment tools, which have used the other of two possible binary outcomes (presence of reoffending). Given that a main objective of the current study was to evaluate the predictive validity of a modified risk assessment tool, the Static-99, and a modified strengths-based tool, the SAPROF, the binary outcome (presence or absence of a new offense during the follow-up period) approach to recidivism versus desistance is adopted. Beyond operationalizations of desistance, an understanding of research design used in studies assessing the predictive validity of forensic risk assessment tools is necessary in order to evaluate the significance of previous findings as well as contextualize the methods employed in the present study.

### **Research Design in Forensic Clinical Assessment**

There are two main research designs used to evaluate forensic assessment tools. The first and most frequently used is retrospective in that it involves archival file coding, typically done by researchers and research assistants trained to use the tool of interest (Viljoen et al., 2012). Importantly, the coding of tools is undertaken blind to eventual outcomes and the coding of outcomes (new offenses occurring over a follow-up period that starts for each individual in the sample upon release) from criminal records is undertaken after coding of the tools has been completed. One major advantage of studies using archived case files is that it is possible to obtain a sample for which the follow-up time is longer than might be possible in a prospective follow-up study. Certainly, whichever design is used, it is crucial that the follow-up is of sufficient length to ensure the baseline of reoffending is adequate for statistical purposes; for studies that include sexual offending as an outcome, the minimal follow-up period is generally considered to be 5 years (Hanson & Bussiere, 1998).

For example, Langton colleagues (2007a) compared the predictive accuracy of six actuarial risk assessment tools: the Static-99, Static-2002, MnSOST, RRASOR, SORAG, and VRAG on a sample of 468 individuals who had sexually offended using a retrospective design. Participants files from the Warkworth Sexual Behaviour Clinic, the Correctional Service of Canada, and the Royal Canadian Mounted Police were reviewed, and the six assessment tools were scored based on the information available in these files. The average follow-up period was 5.9 years, with recidivism information coded from the participants' Canadian Police Information Centre records (Langton et al., 2007a). Langton and colleagues (2007b) used a similar retrospective design to examine the psychometric properties of one actuarial tool, the Static-2002 on the same sample of individuals who had sexually offended.

The second frequently used design is prospective, and involves real-world assessments completed by clinicians in actual cases (Viljoen et al., 2012). For example, Worling and colleagues (2011) used a prospective design, with an average follow-up period of 3.66 years, to assess the validity of the ERASOR among a sample of 191 adolescent males who sexually offended. Participants were assessed at one of five possible agencies by one of 22 clinicians. Each assessment consisted of clinical interviews with both the participant and informants, such as parents, welfare workers, or probation officers, psychological testing, and review of various documents. The clinicians made the ERASOR ratings used in the study after completing these assessments (Worling et al., 2011).

Thus, the two types of design commonly used to evaluate forensic risk assessment instruments, retrospective and prospective, differ with regards to who whether data from individuals who are currently being assessed or the files of participants who were assessed in the past are used. Additionally, they differ in terms of who codes the instruments, with researchers

typically completing the scoring in retrospective designs and clinicians completing scoring in prospective designs. In some cases, there is evidence to suggest that researchers tend to give higher scores than clinicians on forensic assessment tools, such as the Static-99 (Quesada et al., 2014). Given this, knowledge of which type of design was used, and therefore who coded the assessment tool, may aid in contextualizing and interpreting results. Specifically, if researchers give higher scores, or categorize individuals as higher risk than a clinician would, it is possible that the predictive accuracy indicated by a retrospective study is different than might be indicated by a prospective study on the same population. Finally, it should be noted that in the present study a retrospective design was used, with research assistants coding participants' archival files.

### ***Statistical Methods***

In addition to an understanding of research design, comprehension of the common statistical techniques used in this area of research is necessary in order to evaluate the findings of the reviewed studies and provide a basis for their use in the present study. Receiver operating characteristic (ROC) analysis is an index of predictive accuracy frequently used in recidivism research (Guyton & Jackson, 2008). It is well suited to this area of research given that it is less dependent on the base rate of the criterion variable, in this case, sexual recidivism (Douglas et al., 1999). Given that the base rates of sexual recidivism are generally low, techniques which are best suited for use with higher base rates of the criterion variable, approaching 50%, are not as appropriate for use in recidivism research (Douglas et al., 1999).

In the case of risk assessment tools, the ROC curve plots the tool's false alarm rate on the *x*-axis and true positive rate on the *y*-axis (Guyton & Jackson, 2008 Rice & Harris, 1995). The area under the ROC curve (AUC) represents the probability that a randomly selected person who reoffended will score higher on the measure than a randomly selected person who did not

reoffend and can range from zero (100% incorrect prediction) to one (100% perfect prediction). For example, a risk assessment tool that predicts recidivism with perfect accuracy would have an AUC of 1.0. When AUC is equal to 0.5, the tool is as accurate as chance (Douglas et al., 1999; Guyton & Jackson, 2008; Mossman, 1994; Singh, 2012). As a rule of thumb, AUC values less than or equal to 0.63 are considered to have a small effect size, between 0.64 and 0.71 are considered moderate, and equal to or greater than 0.72 are considered a large effect size (Rice & Harris, 2005). The use of ROC analysis allows for researchers to investigate whether a given risk assessment tool, such as the Static-99, predicts recidivism in a given sample with significantly more accuracy than chance. Given that the present study aims to investigate not only whether the Static-99 and SAPROF significantly predict recidivism and desistance, respectively, another statistical technique which allows for the determination of whether one tool possesses incremental validity over another is needed. One such technique is Cox regression, a specific form of survival analysis.

Survival analysis is a group of statistical methods used for analysis of duration data, or time related outcomes, focusing on the time between two events, also known as survival time. Cox regression is a form of survival analysis that is comparable to linear regression; however, unlike linear regression, Cox regression is used to predict the rate of event occurrence (Wright, 2000). In Cox regression, the hazard refers to the risk the event interest (in the case of the present study, sexual or any recidivism) will occur in the immediate future. The hazard function is the dependent variable in Cox regression and can be obtained by plotting the hazard over time (Wright, 2000). Cox regression allows researchers to determine whether higher scores on a given risk assessment tool, such as the Static-99, are associated with a greater likelihood of reoffending.

Iterative Cox regression allows researchers to determine whether scores on one tool, such as the SAPROF, have incremental validity over another tool, such as the Static-99 for the prediction of recidivism outcomes (Wright, 2000; Langton & Worling, 2015). In this form of Cox regression, the first tool (in the present study, the Static-99) is entered in the first step and the tool hypothesized to have incremental validity (in the present study, the SAPROF) is entered in the second step (Wright, 2000; Langton & Worling, 2015). If the second tool does have incremental validity over the first, then it will account for a significant portion of the variance in recidivism and adding it to the model will lead to a significant increase in the amount of variance in recidivism explain by the model as a whole (Langton & Worling, 2015).

### **Demographic Considerations in Risk Assessment**

As previously noted, one of the aims of the presents study was to determine whether the Static-99 and SAPROF, which are intended for use with adults, have predictive validity when used with another age group, adolescents. Careful consideration of the population for which a tool has been developed and from which samples have been used to establish predictive validity is important for a number of reasons, some of which have already been briefly noted. First, as has been previously discussed, most tools are designed for use with a specific population and/or to assess the likelihood of specific types of reoffending. For example, the Static-99 is intended for use with adults who have sexually offended, in order to assess risk for sexual reoffending (Hanson & Thornton, 2000) and its predictive validity for both sexual and violent recidivism for samples from that population has been demonstrated (Barbaree et al., 2001; Beauregard & Mieczkowski, 2009; Langton et al., 2007a; Looman et al., 2013; Phenix & Epperson, 2016; Singh et al., 2011; Stadtland et al., 2005; Swinburne Romine et al., 2012). The use of an assessment tool that has not been tested and demonstrated to have predictive validity for a

specific population (or one for which predictive validity for that population has been tested and has not been demonstrated) would be unethical.

According to the American Psychological Association's (APA; 2013) Specialty Guidelines for Forensic Psychology, clinicians practicing forensic psychology must strive to be accurate and fair (APA, 2013, Section 1.02, 11.01). Further, the Canadian Psychological Association (CPA; 2017) Code of Ethics indicates that clinicians must select assessment tools which are relevant and suited to the characteristics of the individual and based on the best evidence available (CPA, 2017, Section II.18). Given this, it would be an ethical violation to select a tool that was not developed for assessing likelihood of recidivism in the relevant population or shown through research to be a valid predictor of reoffending or desistance when used with that population. One consequence of this is that prospective investigations of such tools with samples from populations for which they were not designed and validated would not be ethical, so studies using archived files would be necessary. Clinicians are also required to promote the well-being and avoid doing harm to their clients and others. Given that inappropriate allocation of services resulting from reliance on the assessments with such tools may cause harm to a client, either by being more restrictive than is necessary or not adequately supportive, and may cause harm to others, if the client needs more intensive care but does not receive it and reoffends, the appropriate selection of validated assessment tools in order to ensure accurate determination of an individual's level of risk and needs is imperative (CPA, 2017, Section II.1 and II.2).

The importance of using tools appropriate for the population being assessed is particularly relevant with regards to age. Almost all tools reviewed above are specifically indicated for use with specific age groups. As discussed, it would be ethically unsound to use

forensic risk assessment tools with populations for which they are not intended (APA, 2013, Section 1.02, 11.01; CPA, 2017, Section II.1, II.2, II.18). In addition, when using tools for an age group they are not intended for use with, some items may not be developmentally sensitive or appropriate for the client's age. For example, the Static-99 age item, which is coded as a risk factor everyone below the age of 25 (Hanson & Thornton, 2000) is inappropriate for use with a youth sample, as the entire sample would have a higher Static-99 score than a sample of adults with the same risk factors present (an issue too in terms of research because scores on that item would afford no variance). Recidivism risk has been shown to change across the lifespan (Barbaree et al., 2009; Hanson, 2002). Given this, the selection of assessment tools intended for use with specified age groups and the omission of items (such as the dichotomous age item on the Static-99) that are developmentally inappropriate on such tools when being investigated by researchers is necessary.

### **The Use of The Static-99 and SAPROF in Adolescent Samples**

As noted above, there is a clear need to attend to sexual offending perpetrated by adolescents (Allen & Superle, 2016; Moreau, 2019), including recidivism. Further, there is a need to incorporate strengths-based tools into risk assessments, ascertain whether such tools have incremental validity over risk-based measures (de Vries Robbé et al., 2011, 2013, 2015b), and assess whether a given measure is suitable for use with adolescents before clinicians begin to use it (APA, 2013, Section 1.02, 11.01; Barbaree et al., 2009; CPA, 2017, Section II.1, II.2, II.18; Hanson, 2002; Hanson & Thornton, 2000). Given this, the present study investigated a modified version of two measures: the Static-99, an actuarial, risk-based tool, and the structured professional judgement, strengths-based SAPROF with a sample of adolescents who have sexually offended. The aim of this study was to assess whether scores on these tools are valid

predictors of two categories of reoffense (or desistance from offending in those two categories) in an adolescent sample, and whether the modified SAPROF has incremental validity over the modified Static-99 in predicting these outcomes.

## **The Static-99**

### *Overview and Development*

As previously mentioned, the Static-99 was created as a combination of two measures intended to assess recidivism risk in individuals who have sexually offended: the Rapid Risk Assessment for Sex Offense Recidivism (RRASOR; Hanson, 1997) and the Structured Anchored Clinical Judgement (SACJ-Min; Grubin, 1998). The RRASOR consists of a small set of variables used to assess recidivism risk: prior sex offenses, male victims, unrelated victims, and age below 25 (Hanson, 1997). The SACJ-Min is used to categorize individuals into risk categories based on their convictions (current sexual offenses, previous nonsexual violent offenses, and four or more previous sentences), and then adjusts their classification based on specific risk factors (stranger victims, male victims, never married, noncontact sexual offenses, residential care in childhood, psychopathy, substance abuse, and deviant sexual arousal) (Grubin, 1998). Hanson and Thornton (2000) found that the two scales combined predicted sexual recidivism more accurately than either scale alone, and thus developed the Static-99 by combining them.

The Static-99 is an actuarial measure of long-term recidivism risk intended for use with adults who have sexually offended. It consists of 10 items, which can be viewed in Table 1. Nine of the items are scored either as yes (one) or no (zero). It is important to note that one of these items, age above or below 25, was revised in the Static-99R, with age coded as one of four categories rather than dichotomously (Helmus et al., 2012). One item, prior sex offenses, is

scored on a scale from zero to three based on the number of charges and convictions. The items are summed in order to create a total score. An individual's score represents their static risk, and thus cannot be used to measure change with treatment (Hanson & Thornton, 2000). Notably, the Static-99 is not intended to be used as a comprehensive assessment tool, as it does not assess dynamic factors or other potentially relevant domains (Hanson & Thornton, 2000).

**Table 1.***Static-99R Items*

| <b>Item</b>                                                | <b>Code (Score)</b>                                                            |
|------------------------------------------------------------|--------------------------------------------------------------------------------|
| 1. Age at release from index sex offense                   | 18-34.9 years (1)<br>35-39.9 years (0)<br>40-59.9 years (-1)<br>60+ years (-3) |
| 2. Single – Ever lived with a lover for at least two years | No (1)<br>Yes (0)                                                              |
| 3. Index nonsexual violence                                | Yes (1)<br>No (0)                                                              |
| 4. Prior nonsexual violence                                | Yes (1)<br>No (0)                                                              |
| 5. (a) Prior sex offenses – charges                        | None (0)<br>1-2 (1)<br>3-5 (2)<br>6+ (3)                                       |
| 5. (b) Prior sex offenses – convictions                    | None (0)<br>1 (1)<br>2-3 (2)<br>4+ (3)                                         |
| 6. Prior sentencing dates (excluding index)                | 3 or less (0)<br>4 or more (1)                                                 |
| 7. Any convictions for noncontact sex offenses             | Yes (1)<br>No (0)                                                              |
| 8. Any unrelated victims                                   | Yes (1)<br>No (0)                                                              |
| 9. Any stranger victims                                    | Yes (1)<br>No (0)                                                              |
| 10. Any male victims                                       | Yes (1)<br>No (0)                                                              |

*Use with Adults***Predictive Validity**

The Static-99 is widely used as a risk assessment tool for adults who have sexually offended, and its predictive validity and other psychometric properties have been extensively examined in the literature (Singh et al., 2011; Tully et al., 2013). Across numerous studies, the

Static-99 has been found to significantly predict recidivism (Barbaree et al., 2001; Beauregard & Mieczkowski, 2009; Langton et al., 2007a; Looman et al., 2013; Phenix & Epperson, 2016; Stadtland et al., 2005; Swinburne Romine et al., 2012). In a meta-analysis of 43 studies, it was found to have moderate predictive validity overall for sexual recidivism when used with adult males who have sexually offended (mean AUC =.69); however, there was notable variation in predictive validity between studies in the analysis (Tully et al., 2013). Another meta-analysis examined violent risk assessment tools, including the Static-99 across 88 independent samples from 68 studies (Singh et al., 2011). This study included a total of 25,980 participants with an average age of 31 years, and who were predominantly male. Approximately half of the included studies used a prospective research design, whereas the other half used retrospective file coding. They found that the Static-99 had moderate predictive validity for violent recidivism risk (mean AUC =.70). Of note, there were no significant differences in the effect sizes of studies using retrospective versus perspective design (Singh et al., 2011).

Evidence for the predictive validity of the Static-99 has been found for a number of types of reoffending (Barbaree et al., 2001; Langton et al., 2007a; Stadtland et al., 2015; Swinburne Romine, 2012; Turner et al., 2016). Swinburne Romine (2012) investigated the predictive validity of the Static-99 in a sample of adults enrolled in an outpatient treatment program for individuals who have sexually offended, and found evidence that the Static-99 significantly predicts contact and non-contact sexual reoffending (Swinburne Romine et al., 2012). In another study, among a sample of 73 adults who sexually offended and were either in a prison-based therapeutic unit, had dropped out of treatment, or were assessed in a psychiatric hospital after being accused of sexually offending, the Static-99 was found to have better predictive validity for violent (AUC =.73), non-contact sexual (AUC =.74), and contact sexual reoffending (AUC

=.64) than other widely used risk assessment tools, such as the HCR-20 (Stadtland et al., 2015). Langton and colleagues (2007a) found that, among a sample of 468 individuals who sexually offended, after an average follow up period of 5.9 years, the Static-99 significantly predicted non-violent (AUC =.65), serious (AUC =.64), sexual (AUC =.64), or general recidivism (AUC =.65). Finally, Turner and colleagues (2016) found that, in a sample of 277 individuals convicted of perpetrating sexual offenses against children, the Static-99 significantly predicted general (AUC =.72), violent (AUC= .74) and sexual recidivism (AUC =.83) (Turner et al., 2016). Together, these findings strongly indicate that the Static-99 significantly predicts several forms of recidivism among adults who have sexually offended.

Although there is evidence to suggest that the Static-99 can predict a number of forms of recidivism (Barbaree et al., 2001; Langton et al., 2007a; Stadtland et al., 2015; Swinburne Romine, 2012; Turner et al., 2016), there is some evidence to suggest that the Static-99 is not equally predictive for all types of individuals who have offended. For example, in a meta-analysis examining the effectiveness of various risk assessment tools for individuals who have sexually offended, findings were mixed regarding the predictive accuracy of the Static-99 with specific types of individuals who have offended (Tully et al., 2013). More specifically, there were mixed findings regarding whether it is more effective at predicting reoffense among those convicted of child molestation or rape, as well as those who have committed contact or non-contact offenses (Tully et al., 2013). Another study found that, among a sample of 186 individuals who have sexually offended, after a follow-up period of approximately five years, the Static-99 significantly predicted all forms of recidivism (sexual, violent, and general) for individuals who had molested victims outside of their family (AUC =.72) and those who had committed incest-related offenses (AUC =.68) (Bartosh et al., 2003). Conversely, it did not

significantly predict any or specifically sexual recidivism among individuals who had committed rape, or non-contact sexual offenses (Bartosh et al., 2003). Together, these findings indicate that the type of sexual offense an individual has committed may impact the predictive validity of the Static-99; however, it is unclear which types of offenses are associated with better or worse performance of the Static-99.

The ethnicity of participants has also been found to influence the predictive accuracy of the Static-99. Varela and colleagues (2013) found evidence for the predictive validity of the Static-99; however, its predictive accuracy was lower for Latino (AUC =.54) participants who have sexually offended than Black (AUC =.65) and White (AUC =.63) participants. Conversely, another study failed to replicate this finding, with no evidence indicating the Static-99 is less predictive of recidivism for Latino (AUC =.64) than White (AUC =.64) or Black (AUC =.62) individuals who have sexually offended (Boccaccini et al., 2017). Of particular relevance to Canadian research, Indigenous peoples who have offended tend to have significantly higher total scores on the Static-99 (Babchishin et al., 2012). Despite this, the tool has been found to significantly predict sexual recidivism among Indigenous peoples who have sexually offended (AUC =.69), and the predictive accuracy of the total score has not been found to vary significantly between Indigenous and non-Indigenous samples (Babchishin et al., 2012). Similarly, Lee and colleagues (2019) found that the Static-99R successfully discriminated between individuals who reoffended and those who did not for both Indigenous (AUC =.66) and White samples (AUC =.72). Thus, while there is some evidence to suggest that the Static-99 is not an accurate predictor of recidivism across all samples (Varela et al., 2013), there is also plenty of evidence to suggest that this is not the case (Babchishin et al., 2012; Boccaccini et al., 2017; Lee et al., 2019).

Overall, it is clear that, when used with adult samples, the Static-99 is a valid predictor of recidivism (Barbaree et al., 2001; Beauregard & Mieczkowski, 2009; Langton et al., 2007a; Looman et al., 2013; Phenix & Epperson, 2016; Stadtland et al., 2005; Swinburne Romine et al., 2012). Despite this, predictive validity is not the only psychometric property that must be considered when determining whether a measure is suitable for assessing reoffending risk in practice. Interrater reliability must also be considered, as a risk assessment tool which indicates different levels of risk for the same individual, depending on who has conducted the assessment, may result in inaccurate or unfair decisions regarding the individual's likelihood of reoffending, and thus may also lead to inappropriate decisions regarding level of care.

### **Interrater Reliability**

As just noted, it is important to consider inter-rater reliability when determining whether a tool is appropriate for use with a given population. Fleiss (1986) proposed intra-class correlation coefficient (ICC) critical values levels to be used to determine the strength of interrater reliability. ICC values greater than or equal to .75 are considered excellent, from .60 to .75 are considered good, from .40 to .60 are considered moderate, and less than .40 are considered poor (Fleiss, 1986). Evidence of good interrater reliability has generally been reported for the Static-99 (Boccaccini et al., 2012; Langton et al., 2007a; Quesada et al., 2014; Storey et al., 2012). Langton and colleagues (2007a) reported adequate inter-rater reliability for the Static-99 based on two independent rater's coding of a subset of 25 cases out of the total sample. Similarly, another study found excellent interrater agreement on Static-99 scores for two samples of adults who have offended, with ICC values of .79 for the first sample and .88 for the second (Boccaccini et al., 2012).

While many studies report the interrater reliability of the Static-99 among researchers, some studies have investigated agreement between clinicians and researchers, as these are the two major groups frequently using the measure. In one study, agreement of Static-99 scores between clinicians and researchers was investigated and found to vary in agreement depending on the item, ranging from moderate (ICC = .56) to excellent (ICC = .89). (Storey et al., 2012). Despite this variation, both clinicians' and researchers' scores moderately predicted sexually violent recidivism (Storey et al., 2012). In another study, it was found that in approximately half of cases researchers and clinicians gave comparable scores, and that researchers tended to give higher scores overall (Quesada et al., 2014). Altogether, these findings indicate that there is some agreement between researchers and clinicians on Static-99 scores; however, there are some differences, with researchers often giving higher scores (Quesada et al., 2014).

One possible explanation for this difference lies in the amount of information that is typically available to clinicians, as compared to researchers. Specifically, clinicians may conduct an assessment a notable amount of time before the individual is released into the community, and thus in the time between assessment and release an individual's risk level may change due to shifts in dynamic factors. Conversely, researchers typically have access to complete files, and thus their assessments are based on the most up-to-date information. Another possible explanation is that researchers are typically trained to strictly adhere to coding guidelines and participate in coding meetings to ensure that all researchers working on a given project are coding the files reliably. Clinicians do not have such controls to prevent them from deviating from the coding guidelines to some extent.

Taken altogether, these findings indicate that the Static-99 is a reliable and valid risk assessment tool for use with adults who have sexually offended (Barbaree et al., 2001;

Beauregard & Mieczkowski, 2009; Langton et al., 2007a; Looman et al., 2013; Phenix & Epperson, 2016; Stadtland et al., 2005; Swinburne Romine et al., 2012). This includes sexual recidivism (Barbaree et al., 2001; Langton et al., 2007a; Phenix & Epperson, 2016; Stadtland et al., 2005; Swinburne Romine et al., 2012; Tully et al., 2013), and there is also some evidence to suggest that it significantly predicts other forms of recidivism, such as violent or general (Barbaree et al., 2001; Langton et al., 2007a; Singh et al., 2011; Stadtland et al., 2005).

Individual characteristics, such as race and type of offense committed may impact the predictive validity of the Static-99 (Bartosh et al., 2003; Tully et al., 2013; Varela et al., 2013); however, there is evidence to suggest that the Static-99 is robust to these idiosyncrasies, particularly those relating to ethnicity (Babchishin et al., 2012; Boccaccini et al., 2017; Lee et al., 2019). Finally, the Static-99 appears to have good interrater reliability (Boccaccini et al., 2012; Langton et al., 2007a), although there may be some disagreement between scores assigned by clinicians versus researchers (Quesada et al., 2014). Given these findings, the Static-99 may be considered appropriate for use with adults who have sexually offended; however, in order to justify use with other populations, such as with adolescents, evidence for predictive validity and reliability of this measure must be demonstrated for those populations. Further, the Static-99's clearly demonstrated strong psychometric properties when used with adults indicate that such an investigation is worth undertaking.

### ***Use with Adolescents***

As previously noted, the Static-99 was designed for use with adult populations and there are potentially several problems associated with its use with other age groups. Two Static-99 items would not vary for individuals below the age of 18, effectively rendering it an 8-item tool if used with adolescents. These include the aforementioned age and relationship status items. If

the invariant scores on these items with adolescents is not an accurate reflection of increased risk for that group, use of the Static-99 to assess risk of re-offense in adolescents who have sexually offended is unethical, as discussed above (APA, 2013, Section 1.02, 11.01; CPA, 2017, Section II.1, II.2, II.18). Finally, there are risk assessment measures specifically intended for use with adolescents, so the use of a measure that has not been adequately researched for that population would represent ethically unsound practice (but, importantly, would underscore the potential value of research, using an archival coding approach, to evaluate tools such as the Static-99, which have represented such influential gains in the field).

As discussed above, the Juvenile Sex Offender Assessment Protocol (J-SOAP-II; Prentky & Righthand, 2003) and the Juvenile Sexual Offense Recidivism Risk Assessment Tool - II (JSORRAT-II; Epperson et al., 2006) are both empirically validated risk assessment tools for use with adolescents with sexual offenses. Evidence for both predictive validity for sexual and general recidivism, and interrater reliability has been reported for the J-SOAP-II when used with adolescents who have sexually offended (Prentky & Righthand, 2003; Viljoen et al., 2012). The JSORRAT-II has been found to significantly predict sexual recidivism in adolescents who have sexually offended (Epperson & Ralston, 2015; Ralston et al., 2016; Viljoen et al., 2012) and have high interrater reliability (Ralston et al., 2016). Given that there are tools available and shown to predict sexual reoffending among adolescents who have sexually offended, and ethical and practical difficulties associated with the use of the Static-99 with adolescents, validation of the Static-99 in adolescent samples is necessary in order to justify its use.

Despite having been developed for use with adult populations, a small number of studies have examined the efficacy of the Static-99 for predicting reoffending among adolescents who have sexually offended. See Table 2 for a summary of key findings with regards to the use of the

Static-99 in adolescent samples. Viljoen and colleagues (2012) conducted a meta-analysis of 33 studies in order to examine the predictive validity of several sexual recidivism risk tools, including the Static-99, J-SOAP-II, and JSORRAT-II when used to predict sexual reoffending among adolescents who have sexually offended. Eight of the studies included were conducted with the Static-99. They found that the Static-99 did predict sexual reoffending among adolescents who had sexually offended ( $AUC = .67$ ), and that measures designed specifically for use with adolescents, such as the JSORRAT-II, did not perform significantly better than the Static-99. One study included in this analysis, Viljoen and colleagues (2009), did not find support for the predictive validity of the Static-99 when used with adolescents.

Other studies have also found evidence of its predictive validity when used with adolescents. Ralston and Epperson (2013) specifically examined the predictive validity of risk assessment instruments intended for use with adults, including the Static-99, among a sample of adolescents who have sexually offended. The sample consisted of 636 males and the risk assessment instruments were retrospectively coded based on case files obtained from Utah's Juvenile Court and Juvenile Justice Services. They found that the Static-99's predictive accuracy for adolescent samples ( $AUC = .72$ ) was comparable to its predictive accuracy for adults (Ralston & Epperson, 2013). Further, as was found by Viljoen and colleagues (2012), the predictive accuracy of the Static-99 was similar to that of measures developed for use with adolescents (Ralston & Epperson, 2013). Finally, Poole and colleagues (2000) examined the predictive power of the Static-99 with a sample of males who sexually offended before the age of 18, and found that the total score was significantly correlated with recidivism. It should be noted that this finding should be interpreted with caution as, given the low base rates of sexual recidivism, correlation is not the best suited method for assessing predictive validity (Douglas et al., 1999).

**Table 2.***Key Findings with Regards to the Use of the Static-99 in Adolescent Samples*

|                          | <b>Sample</b>                                                                                                     | <b>Age of Participants</b>            | <b>Omitted Items</b>     | <b>Operationalization of Recidivism</b>        | <b>Key Findings</b>                                                                                                                                                                                                                                                                           |
|--------------------------|-------------------------------------------------------------------------------------------------------------------|---------------------------------------|--------------------------|------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Poole et al., 2000       | Males who committed a sexual offense below the age of 18 (N = 49)                                                 | Mean age of participants not reported | None                     | Arrest for a new sexual offense                | I. Static-99 total score was significantly correlated with recidivism                                                                                                                                                                                                                         |
| Ralston & Epperson, 2013 | Male juveniles who have sexually offended (N = 636)                                                               | 15.2 years                            | None                     | New criminal charges                           | I. Static-99 significantly predicted recidivism in the short term<br>II. Static-99's predictive accuracy with juveniles was comparable to its accuracy with adults<br>III. Static-99's predictive accuracy was similar to that of measures intended for juveniles (MnSOST-R, JRAS, J-SOAP-II) |
| Viljoen et al., 2009     | Male adolescents enrolled in a residential treatment program for individuals who have sexually offended (N = 193) | 15.3 years                            | "Young age" item omitted | New criminal charges                           | I. Predictive validity of the Static-99 did not significantly vary by age<br>II. Static-99 did not significantly predict sexual or nonsexual reoffending                                                                                                                                      |
| Viljoen et al., 2012     | Adolescents who committed sexual offences (K = 33; N = 6196)                                                      | 16 years                              | Not applicable           | Mixed: new criminal charges or new convictions | I. Static-99 significantly predicted sexual reoffending<br>II. Tools designed for use with adolescents (J-SOAP-II, ERASOR, and JSORRAT-II) were not better predictors of sexual offending than the Static-99                                                                                  |

Note: Viljoen et al. (2012) is a meta-analysis of 33 studies, 20 of which were unpublished and 8 of which included the Static-99.

- a. Juvenile Sex Offender Assessment Protocol-II (J-SOAP-II; Prentky & Righthand, 2003)
- b. Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR; Worling & Curwen, 2001)
- c. Juvenile Sexual Offense Recidivism Risk Assessment Tool-II (JSORRAT-II; Epperson et al., 2006)
- d. Minnesota Sex Offender Screening Tool-Revised (MnSOST-R; Epperson et al., 2004)
- e. Juvenile Risk Assessment (JRAS; Hiscox et al., 2007)

In conclusion, although there is evidence to support the use of the Static-99 with adult populations (Barbaree et al., 2001; Langton et al., 2007a; Phenix & Epperson, 2016; Singh et al., 2011; Stadtland et al., 2005; Swinburne Romine et al., 2012; Tully et al., 2013), findings regarding its validity when used with adolescent populations have been mixed (Ralston & Epperson, 2013; Viljoen et al., 2009; Viljoen et al., 2012). Given the developmental and ethical concerns associated with using an adult measure with adolescent populations, previously described, further research is needed in order to determine if the Static-99 is a valid predictor of recidivism for adolescents and thus suitable for use with adolescent populations. Since the Static-99 is a widely used measure with clearly demonstrated predictive validity in adults, investigation of its predictive validity with adolescents is clearly warranted.

## **The SAPROF**

### ***Overview and Development***

As described above, the SAPROF is a structured professional judgement strengths-based or assessment tool which assesses primarily dynamic protective factors. It is intended for use in conjunction with risk assessment measures, to estimate the likelihood that individuals who have committed violent offenses and patients in forensic psychiatric settings will engage in future violent behaviour (de Vogel et al., 2011). It has three conceptual subscales: internal, with items such as empathy and intelligence; motivational, with items such as work and medication; and external, with items such as living circumstances and professional care (de Vogel et al., 2011). See Table 3 for a list of SAPROF items. Each item is coded on a three-point scale: zero for does not apply, one for probably applies, and two for definitely applies. After the items are coded, critical items can be identified; these are items that the assessor deems important in the individual case for focusing intervention efforts to prevent that person's perpetration of future

violence. Items can then be categorized as key items, which are already present, or goal items, which may be present after intervention. A final protection judgement is made, in which the assessor gives a summary rating of a low, moderate, or high level of protection based on their scoring of the items and their own integration of the information. Lastly, the results of the SAPROF are integrated with a risk based SPJ tool, such as the HCR-20, to provide a final, overall estimation of risk for future violent behaviour (de Vogel et al., 2011).

**Table 3.***SAPROF Items*

| <b>Subscale</b>       | <b>Item</b>                                                                                                                                                       |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Internal</i>       | 1. Intelligence<br>2. Secure attachment in childhood<br>3. Empathy<br>4. Coping<br>5. Self-control                                                                |
| <i>Motivational</i>   | 6. Work<br>7. Leisure activities<br>8. Financial management<br>9. Motivation for treatment<br>10. Attitudes towards authority<br>11. Life goals<br>12. Medication |
| <i>External Items</i> | 13. Social network<br>14. Intimate relationship<br>15. Professional care<br>16. Living circumstances<br>17. External control                                      |

The SAPROF was developed via the selection of protective factors which have been shown empirically to be related to desistance from violent behaviour (de Vogel et al., 2011). Further, the SAPROF was created with the intention of assessing dynamic protective factors, as this allows for those factors to be targeted in treatment. This tool was originally intended for use with adult males with a history of perpetrating sexual violence, who also met diagnostic criteria for a psychiatric disorder (de Vogel et al., 2011). The developers noted that it could also be used in women who meet the same criteria, with some caution. Given a lack of empirical support for the use of the SAPROF with women, it may be unethical to do so for the same reasons it is unethical to use tools that have not been shown to accurately predict reoffending in adolescents, with adolescents (APA, 2013, Section 1.02, 11.01; CPA, 2017, Section II.1, II.2, II.18). As

previously noted, the SAPROF was developed with the intention of providing a complementary tool for use with risk assessment measures, thus providing a more complete view of a given individual's context, by structuring the inclusion in an assessment of the assessee's strengths rather than their deficits or risk factors alone (de Vogel et al., 2011).

### *Use with Adults*

#### **Predictive Validity**

There is a growing body of literature indicating that the SAPROF is a valid predictor of desistance (Coupland & Olver, 2020; de Vries Robbé et al., 2011, 2013, 2015b; Kashiwagi et al., 2018). In one study which retrospectively coded the files of 467 incarcerated males who have committed a sexual offense, after an average follow up period of 5.8 years, SAPROF total scores moderately predicted desistance from violent recidivism (AUC =.64) (Coupland & Olver, 2020). Further, they found that the internal and motivational scales had a small but significant predictive effect (AUC =.64 and .66, respectively), and the external subscale did not significantly predict desistance (AUC =.56) (Coupland & Olver, 2020). Another study examined the predictive validity of the SAPROF using retrospective file evaluation in a sample of male patients in forensic psychiatric settings with a history of violent offending (de Vries Robbé et al., 2011). These researchers reported that SAPROF total scores significantly predicted desistance from violent offending one- (AUC =.85), two- (AUC =.80), and three-year follow-ups (AUC =.74). Further, the final protection judgement and integrated final risk judgements of the SAPROF significantly predicted recidivism at one-(AUC =.82 and .80, respectively), two- (AUC =. 77 and .72, respectively) and three-year follow-up (AUC =.71 and .65, respectively) (de Vries Robbé et al., 2011).

de Vries Robbé and colleagues (2015b) investigated the SAPROF's accuracy for predicting desistance from sexually violent reoffending by retrospectively coding the files of 83 adult males who have sexually offended. They found that the SAPROF total score significantly predicted desistance from sexually violent (AUC =.76) and generally violent offending (AUC =.83), and this persisted once scores on risk factor assessment measures were controlled for. Finally, Yoon and colleagues (2018), using retrospective analysis of the files of 467 individuals who have offended sexually, found that the SAPROF total score had moderate predictive accuracy for desistance from violent offending (AUC =.61). Further, the internal and motivational subscales significantly predicted desistance from violent offending (AUC =.57 and .63, respectively), and, as was found by Coupland and Olver (2020), the external scores did not predict desistance (AUC =.54) (Yoon et al., 2018). Overall, these findings indicate that the SAPROF is a valid predictor of desistance (Coupland & Olver, 2020; de Vries Robbé et al., 2011, 2013, 2015b; Kashiwagi et al., 2018).

Conversely, some studies have found evidence indicating that the SAPROF is not a valid predictor of desistance from violent behaviour in all populations. Although the authors of the SAPROF have indicated that it can be used with women, they suggest doing so should be done with caution; in one of the developers' studies the SAPROF did not significantly predict desistance from general (AUC =.39) or violent reoffending (AUC =.50) in a sample of female patients in a forensic psychiatric setting (de Vogel et al., 2019). Further, Viljoen and colleagues (2016) compared the predictive validity of the SAPROF across genders. A prospective design was used, with the files of 102 patients in a psychiatric hospital reviewed in addition to interviews with the participants. Follow up data was collected one year after initial data collection. SAPROF total scores were found to have a non-significant, positive correlation with

severe aggression in female psychiatric patients, indicating that the SAPROF may not only not significantly predict desistance in female samples, it actually may be somewhat associated with greater risk for recidivism (Viljoen et al., 2016). This may indicate that the protective factors included in the SAPROF are not protective for females, or alternatively, that some of the protective factors of the SAPROF are actually risk factors for desistance in female populations.

The SAPROF has also been investigated as a predictor of outcomes other than desistance from criminal offending. For example, Abdin and colleagues (2013) examined its ability to predict violence and self-injury in inpatient forensic psychiatric settings. They found that the SAPROF predicted both the absence of violence (AUC =.84) and of self-harm (AUC =.76). Specifically, the empathy, coping ability, self-control, financial management, motivation for treatment, work and leisure, and attitudes towards authority items predicted absence of violence. The coping, empathy, leisure activities, self-control, and use of medication items predicted the absence of self-harm (Abdin et al., 2013). These findings indicate that the SAPROF can be used to predict more outcomes than just desistance from violence, despite its intended use. Similarly, another study examined inpatient aggression among a predominantly male sample of patients in forensic psychiatric settings who had committed violent or sexual offenses. The SAPROF had good predictive validity for a lack of inpatient aggression (AUC =.75), which included physical or verbal aggression that resulted in the patient being confined in a recovery room or receiving criminal charges (de Vries Robbé et al., 2016). Persson and colleagues (2017) corroborated these results in a sample of individuals who were detained while undergoing forensic psychiatric investigations, finding that the SAPROF significantly predicted a lack of violent behaviour during imprisonment (AUC =.78).

de Vries Robbé and colleagues (2013) investigated the value of utilizing both risk and protective factor assessments together; in this case, the SAPROF was used as a measure of protective factors and the HCR-20 was used to assess risk factors. Their sample consisted of 188 male patients in a Dutch forensic psychiatric hospital, who had sexually or violently offended. Recidivism included any new conviction for sexual or non-sexual violent offenses after discharge, after follow-up at one and three years, as well as a mean long-term follow-up period of 11 years. They found that the SAPROF total score significantly predicted desistance at one-year (AUC =.85), three-year (AUC =.75), and long term follow up (AUC =.73), and that this remained significant when the authors controlled for HCR-20 scores. Further, they found that the two measures had nearly equal predictive validity at one- and three-year follow-up, but the SAPROF total score's long-term predictive validity was stronger than that of the HCR-20 total score (AUC =.73 and .64, respectively). Notably, the combined HCR-SAPROF index was the best predictor of violent reoffending at one- (AUC =.87) and three-year follow up (AUC =.76) (de Vries Robbé et al., 2013).

This has also been found in other studies, with combined HCR-SAPROF index found to be the best predictor of violent recidivism for a short-term follow-up period (de Vries Robbé et al., 2011). The SAPROF and HCR-20 were strongly, negatively correlated with each other (de Vries Robbé et al., 2013). This finding has been reported across a number of studies (de Vries Robbé et al., 2013, 2015b). This is notable as it indicates that the two measures may simply mirror each other, and thus may not add anything unique to the assessment of recidivism risk. This possibility notwithstanding, the SAPROF was found to have incremental predictive validity over the HCR-20, indicating that adding protective factors increased predictive validity over the use of risk factors alone in that sample (de Vries Robbé et al., 2013). Thus, there is some

evidence to suggest that the SAPROF is a valid predictor of desistance among adults who have offended (Coupland & Olver, 2020; de Vries Robbé et al., 2011, 2013, 2015b; Kashiwagi et al., 2018). Further, there is some evidence to suggest that it may be incrementally valid over risk-based measures, such as the HCR-20 (de Vries Robbé et al., 2011, 2013, 2015b). As with the Static-99, this information is not sufficient to determine whether the SAPROF is suitable for use to assess the likelihood an individual who has offended will desist. In order to ensure that the SAPROF consistently and accurately predicts desistance, evidence of inter-rater reliability is also necessary.

### **Inter-rater Reliability**

Researchers have reported adequate levels of inter-rater reliability for the SAPROF (Persson et al., 2017; de Vries Robbé et al., 2011, 2015b; Yoon et al., 2018). One study found that the total score and final protection judgements had good interrater reliability when used for a sample of male patients in forensic psychiatric settings who have violently offended, based on the independent coding of 40 files from the larger sample by two independent raters. The intra-class correlation coefficient (ICC) was .88 for the total score and .85 for the Final Protection Judgements (de Vries Robbé et al., 2011). de Vries Robbé and colleagues (2015b) replicated this finding, with two independent raters coding a subset of 30 files out of the entire sample. Excellent interrater reliability was found for the SAPROF total score (ICC = .85) and good interrater reliability was found for the final protection judgement (ICC = .73). Persson and colleagues (2017) reported an ICC value of .86 for the SAPROF total score; however, it is unclear how inter-rater reliability was assessed in this study. Finally, Yoon and colleagues (2018) found good interrater reliability (ICC = .72) based on three independent raters coding 30

cases out of the total participants files. Altogether, these findings indicate that the SAPROF has good interrater reliability.

Overall, the SAPROF has been found to have good interrater reliability (Persson et al., 2017; de Vries Robbé et al., 2011, 2015b; Yoon et al., 2018) and be a valid predictor of desistance among adults who have sexually offended (Coupland & Olver, 2020; de Vries Robbé et al., 2011, 2013, 2015b; Kashiwagi et al., 2018); however, some studies have reported conflicting findings, indicating that it is not a significant predictor of desistance (Coid et al., 2015; de Vogel et al., 2019). Finally, there is evidence to suggest that the SAPROF has incremental validity over risk-based measures, such as the HCL-20 (de Vries Robbé et al., 2013). These findings indicate that SAPROF may be suited for use assessing the likelihood of desistance in individuals who have offended; however, further research is needed given conflicting findings regarding predictive validity. Further, although the SAPROF may be suitable for adults, in order to determine its suitability for use with other populations, such as adolescents, evidence for predictive validity and interrater reliability with such populations is needed. Given that the SAPROF is well validated with adults, one of few strengths-based tools, and be incrementally valid over risk-based tools, such an investigation of its predictive validity with adolescents is likely a worthwhile endeavor.

### ***Use with Adolescents***

Although there is evidence supporting the use of the SAPROF for assessing protective factors and predicting recidivism in adult populations, further research is needed to determine if it is suitable for use with adolescents. As with the Static-99, the SAPROF includes some items which are not developmentally appropriate: financial management, intimate relationship, and to some extent, work. As previously described, the presence of developmentally inappropriate items

may skew the scores of adolescent samples, so the previously cited ethical concerns associated with such items are also relevant to the use of the SAPROF with adolescents. Further, there are strengths-based measures available that are specifically intended for use with adolescents, such as the SAPROF-YV, which has been shown to be a valid predictor of reoffending in its target population (Chu et al., 2020; Kleeven et al., 2020), making the use of an instrument that has not been validated in order to make decision which may drastically impact an individual's life particularly questionable. Despite this, given the well-supported predictive validity of both the Static-99 and SAPROF, it may be worthwhile to determine whether either tool also possess predictive validity with adolescent populations. If evidence were found indicating that these measures have comparably high predictive validity when used with adolescents, it would mean that two tools which are known to be accurate predictors of recidivism can be used with adolescents who have sexually offended. This may provide the means for more accurate risk assessment and appropriate judgements regarding treatment and level of support needed.

There is a notable dearth of studies investigating the validity of the adult version of the SAPROF when used with adolescent populations. See Table 4 for a review of key findings. One study investigated the predictive validity of the SAPROF among a sample of 71 male adolescents who had sexually offended using retrospective rating of participants' clinical files (Klein et al., 2015). They found that, after an average follow up period of 47.84 months, the SAPROF total score (AUC =.65) and final protection judgement (AUC =.64) significantly predicted desistance from violent offending. Additionally, they found that the internal subscale significantly predicted desistance (AUC =.68), but the motivational (AUC =.61) and external subscales (AUC =.53) did not (Klein et al., 2015). Zeng and colleagues (2015) examined the predictive validity of the SAPROF when used with a sample of 97 male youth from Singapore who had sexually offended.

The participants' files were retrospectively viewed in order to the SAPROF items to be rated. They found that, after a mean follow-up period of four and a half years, both the total and domain scores of the SAPROF failed to significantly predict desistance from both sexual (AUC =.48) and non-sexual offending (AUC =.62) (Zeng et al., 2015). Given the lack of evaluation studies of the SAPROF with adolescents, and the conflicting findings of the studies that do exist, there is a clear need for further research before any conclusions regarding the SAPROF's predictive validity for adolescents can be made. Thus, the present study aimed to evaluate the predictive validity of the SAPROF among a sample of adolescents who have sexually offended, as well as determine whether the SAPROF has incremental validity over a risk-based tool, the Static-99.

**Table 4.**

*Key Findings with Regards to Use of the Structured Assessment of Protective Factors (SAPROF) in Adolescent Samples*

|                    | <b>Sample</b>                                               | <b>Age of Participants</b> | <b>Omitted Items</b>                                                                                                                                      | <b>Operationalization of Recidivism</b> | <b>Key Findings</b>                                                                                                                                                                |
|--------------------|-------------------------------------------------------------|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Zeng et al., 2015  | Male youth in Singapore who have sexually offended (N = 97) | 15 years                   | None                                                                                                                                                      | New criminal charges                    | I. Total SAPROF scores did not significantly predict sexual or nonsexual recidivism<br>II. SAPROF domain scores did not significantly predict sexual or nonsexual recidivism       |
| Klein et al., 2015 | Male juveniles with an index sex charge (N = 71)            | 14.5 years                 | Life goals and medication items omitted. The financial management item changed to “dealing with pocket money.” Work item changed to “school achievement.” | New criminal charges                    | I. SAPROF total scores were negatively correlated with violent and general recidivism<br>II. SAPROF sum and SAPROF-SPJ scores significantly predicted a lack of violent recidivism |

### **Present Study**

Previous research indicates that the Static-99 is a valid predictor of sexual recidivism among adults who have sexually offended (Barbaree et al., 2001; Beauregard & Mieczkowski, 2009; Langton et al., 2007a; Looman et al., 2013; Phenix & Epperson, 2016; Stadtland et al., 2005; Swinburne Romine et al., 2012), and may also be useful for assessing risk for reoffense in adolescents who have offended (Ralston & Epperson, 2013; Viljoen et al., 2012). The SAPROF has also been widely affirmed as a valid predictor of non-recidivism (Coupland & Olver, 2020; de Vries Robbé et al., 2011, 2013, 2015b; Kashiwagi et al., 2018), with some findings indicating it may also be an effective measure of protective factors in adolescents (Klein et al., 2015).

Further, the SAPROF has been found to have incremental validity over one risk-based

assessment tool, the HCR-20, in adult populations (de Vries Robbé et al., 2013). Given these findings, the present study aimed to investigate the use of a modified version the Static-99 and the SAPROF in a sample of adolescents who have sexually offended to predict serious (violent or sexual) and any recidivism, as well as determine whether the modified SAPROF is incrementally valid over the modified Static-99.

### *Hypotheses*

Based on the reviewed literature, three hypotheses were postulated. The first two concern main effects of the modified Static-99 and SAPROF (that is, their predictive accuracy). The third hypothesis concerns the incremental predictive validity of the modified SAPROF. See Table 5 for a summary of the present study's hypotheses, the bases for these hypotheses, and the data-analytic strategies to be employed to test them.

- 1) It was hypothesized that the total score for the modified Static-99 would be a statistically significant predictor of both recidivism outcomes, a new serious offense and any new offense during the follow-up period, as evidenced by statistically significant AUC values above .50 for the presence of these outcomes.
- 2) It was hypothesized that the total score for the modified SAPROF would be a statistically significant predictor of both desistance outcomes, the absence of a new serious offense and the absence of any new offense during the follow-up period, as evidenced by statistically significant AUC values above .50 for the absence of these outcomes.
- 3) It was hypothesized that the SAPROF internal conceptual subscale would be a statistically significant predictor of both desistance outcomes, the absence of a new serious offense and the absence of any new offense during the follow up period, as indicated by statistically significant AUC values above .50 for the absence of these

outcomes. Given that the motivational and external conceptual subscales have not been found to predict desistance (Klein et al., 2015), no hypotheses were put forward for these subscales.

- 4) It was hypothesized that the total score for the modified SAPROF would have incremental predictive validity, entering a hierarchical regression model as a statistically significant predictor and producing a statistically significant change in the model already containing the modified Static-99 for both recidivism outcomes, a new serious offense and any new offense during the follow-up period.

**Table 5.***Hypotheses, Justification, and Corresponding Data-Analytic Plan*

| <b>Hypothesis</b>                                                                                                                                                                                          | <b>Empirical Basis</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | <b>Data-Analytic Strategy</b>                                                                                                                                                                                                                                           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Higher scores on the modified Static-99 are associated with a greater probability of reoffending.                                                                                                       | A large number of studies have found that the Static-99 is a valid predictor of violent, non-violent, general, and sexual recidivism in adults who have sexually offended (Barbaree et al., 2001; Beauregard & Mieczkowski, 2009; Langton et al., 2007a; Looman et al., 2013; Phenix & Epperson, 2016; Stadtland et al., 2005; Swinburne Romine et al., 2012). Some studies have found evidence that it is also a valid predictor of general recidivism in adolescent samples (Ralston & Epperson, 2013; Viljoen et al., 2012). | Receiver Operating Characteristic analyses were used to generate Area Under the Curve (AUC) values, with two outcomes (a new serious offense and any new offense) as dependant variables.                                                                               |
| 2. Lower total scores on the modified SAPROF are associated with a greater probability of reoffending.                                                                                                     | Several studies have found that the SAPROF is a valid predictor of desistance from general, violent, and sexual offending in adults who have sexually offended (Coupland & Olver, 2020; de Vries Robbé et al., 2011, 2013, 2015b; Kashiwagi et al., 2018). One study found that it is also a valid predictor of general desistance in an adolescent sample (Klein et al., 2015).                                                                                                                                                | Receiver Operating Characteristic analyses were used to generate Area Under the Curve (AUC) values, with two outcomes (absence of a new serious offense and absence of any new offense) as dependant variables.                                                         |
| 3. Lower scores on the SAPROF internal conceptual subscale are associated with a greater probability of reoffending. No hypotheses were put forward for the external or motivational conceptual subscales. | The SAPROF internal subscale and final protection judgement have been found to significantly predict desistance from violent offending in an adolescent sample (Klein et al., 2015). The SAPROF external and motivational subscales have not been found to significantly predict desistance (Klein et al., 2015).                                                                                                                                                                                                               | Receiver Operating Characteristic analyses were used for both conceptual subscales to generate Area Under the Curve (AUC) values, with two outcomes (absence of a new serious offense and absence of any new offense) as dependant variables.                           |
| 4. The modified SAPROF has incremental validity over the modified Static-99.                                                                                                                               | de Vries Robbé and colleagues (2011, 2013, 2015b) found that the SAPROF has incremental validity over another risk assessment measure (the HCR-20) in the prediction of sexual and violent recidivism with a sample of adult forensic psychiatric patients who had sexually or violently offended.                                                                                                                                                                                                                              | A pair of Cox regression survival analyses, one with a new serious offense as the dependant variable, and one with any new offense as the dependant variable, were run with the modified Static-99 entered in the first step and modified SAPROF entered in the second. |

## Method

The material below in the subsections on participants, outcome, and procedure is based on material provided in Langton et al. (2021).

### Participants

The sample consisted of adolescent males with a documented sexual offense ( $N = 323$ ) who had a specialized assessment, including an assessment of risk for reoffending, at one of three sites in an urban area in Southern Ontario from 2003 to 2014. The sites included a private practice, a Ministry program which provided specialized services for adolescents who have sexually offended, and a community agency which also provides specialized services for this population. Some cases were removed due to inadequate file materials for the purpose of coding and/or age at release or assessment above or equal to 19 years old. The decision was taken to focus only on male adolescents because of the very small number of female adolescents who received services at these sites.

The follow-up period for each participant began either the date the assessment was completed or the date of release, whichever occurred last. The average age of participants ( $N = 272$ ) at the beginning of the follow up period was 15.99 years ( $SD = 1.52$ , range = 12.10 to 18.90). Eighty-five percent of participants had no previous convictions for a sexual offense, 14% had one prior conviction, and 1% had two or more previous convictions. Ten percent of participants had at least one prior violent, nonsexual offense. Thirty percent of participants did not have a history of nonviolent offending, 42% had one to four prior nonviolent offenses, and 25% had at least five prior nonviolent offenses. Data on the ethnic origins of 118 participants was available. Of those participants, 36% were of European origin, 19% were of Caribbean origin, 13% were of Asian origin (including West Central, Middle Eastern, South, East, and

South East), 12% were of North American Indigenous origin, 11% were of African origin (including North, South, East, West, and Central), and 9% were of Central, South, and Latin American origin. After an average follow-up period of 4.69 years (SD = 3.36 years), 32% of the sample committed any kind of new offense. After a follow up period of 5.20 years (SD = 3.24 years), 24% of the sample committed a new serious (violent or sexual) offense.

## **Measures**

### ***Static-99 (modified)***

As previously described, the Static-99 is an actuarial measure of static risk factors sexual recidivism, intended for use with adults who have sexually offended. It is a unidimensional risk scale. Initially, it was intended to predict two outcomes: sexual and violent recidivism (Hanson & Thornton, 2000; Harris et al., 2003) but the developers have since recommended it be used only to assess risk of sexual recidivism (Phenix et al., 2016). Although the developers indicated that the Static-99 might be used with caution as part of a comprehensive assessment of risk for recidivism with justice-involved youth (Harris et al., 2003), the developers are now clear that the tool is intended for use only with adults (Phenix et al., 2016). It is used to assess ten risk factors shown to be associated with elevated risk for sexual recidivism (Phenix et al., 2016). Eight of the 10 items were coded resulting in a modified version of the Static-99 (the items concerning age and long-term partners were omitted given that the sample was made up of adolescents not adults). The coding manual used was the 2003 version (Harris et al., 2003) because the 2016 version was not available during the time period the data coding was undertaken. As previously noted, there have been mixed findings regarding the suitability of this measure for use with adolescents (Ralston & Epperson, 2013; Viljoen et al., 2009; Viljoen et al., 2012). For an in-depth description of this tool, see the previous section on the Static-99.

***SAPROF (modified)***

The SAPROF, also previously described, is a measure of dynamic protective factors against recidivism intended for use with adults who have violently offended (de Vogel et al., 2011). The SAPROF is made up of three conceptual subscales: internal, external, and motivational. The internal subscale consists of five items including empathy and intelligence the external subscale consists of five items including living circumstances and professional care, and the motivational subscale includes seven items including work and medication (de Vogel et al., 2011). Three items from the Motivational subscale were coded resulting in a modified version of the SAPROF (the items concerning age and long-term partners were omitted given that the sample was made up of adolescents not adults). As with the Static-99, there have been mixed findings with regards to the suitability of the SAPROF for use with adolescent populations (Klein et al., 2015; Zeng et al., 2015). For a comprehensive review of the SAPROF, see the previous section dedicated to it.

***Outcomes***

Recidivism outcomes were coded based on four sources of information: the Royal Canadian Mounted Police national database of criminal convictions, the Canadian Police Information Centre records, the Ministry of Children and Youth Services case files, and the Ministry of Community Safety and Correctional Services tracking information systems for youth and adolescents who have offended. Follow-up periods and the percentage of adolescents with a new offense for both dependant variables: a new serious offense (which included any new contact sexual offense as well as any new violent offense of a nonsexual nature) and any new offense (an inclusive outcome) are reported below.

**Procedure**

Ethical clearance was obtained by Dr. Langton from all three sites, the relevant Ministry committees, as well as his institutional affiliations before the larger project began. Archived case files, which included all reports by case officers, social workers, psychologists, schools, police, and the court, were accessed for all participants in the sample. Three senior undergraduate research assistants were trained to code the modified versions of the Static-99 and SAPROF, in addition to other tools and variables included as part of the larger project, over a period of two weeks. The training included didactic instruction, case studies, independent practice, and group discussion of practice cases. Prior to coding the remainder of the files, a set of 23 cases were coded independently in order to determine interrater reliabilities. Supervision sessions were held weekly throughout the duration of the coding work. The contents of each file were coded on a scale from 1 (very complete) to 5 (very incomplete). Files coded as incomplete or very incomplete were removed from the sample as they lacked key information necessary for coding key variables.

## **Data-Analysis**

### ***Hypothesis 1***

*Higher scores on the modified Static-99 are associated with a greater probability of reoffending.*

Receiver Operating Characteristic analyses were used to generate Area Under the Curve (AUC) values, with two outcomes (a new serious offense and any new offense) as dependant variables. As previously noted, ROC analyses are better suited for use in the present study than alternative approaches, such as correlations, because recidivism occurs at a relatively low base rate (Douglas et al., 1999). In the context of hypothesis 1, AUC values can be interpreted as the probability that a randomly selected participant who reoffended has a higher score on the

modified Static-99 than a randomly selected participant who did not reoffend (Douglas et al., 1999; Guyton & Jackson, 2008; Langton et al., 2007a; Mossman, 1994; Singh, 2012).

### ***Hypothesis 2***

*Lower scores on the modified SAPROF are associated with a greater probability of reoffending.*

As for hypothesis 1, Receiver Operating Characteristic analyses were used to generate Area Under the Curve (AUC) values, with two outcomes (absence of a new serious offense and absence of any new offense) as dependant variables. The AUC values can be interpreted as the likelihood that a randomly selected participant who did not reoffend will have a higher score on the modified SAPROF than a randomly selected participant who did reoffend (Douglas et al., 1999; Guyton & Jackson, 2008; Langton et al., 2007a; Mossman, 1994; Singh, 2012).

### ***Hypothesis 3***

*Lower scores on the SAPROF internal conceptual subscale are associated with a greater probability of reoffending.* As for hypothesis 1 and 2, Receiver Operating Characteristic analyses were used to generate Area Under the Curve (AUC) values, with two outcomes (absence of a new serious offense and absence of any new offense) as dependant variables. The AUC values can be interpreted as the likelihood that a randomly selected participant who did not reoffend will have a higher score on the internal subscale than a randomly selected participant who did reoffend (Douglas et al., 1999; Guyton & Jackson, 2008; Langton et al., 2007a; Mossman, 1994; Singh, 2012). The SAPROF modified motivational and external subscales were also tested although there are no a priori hypotheses for either subscale.

### ***Hypothesis 4***

*The modified SAPROF has incremental validity over the modified Static-99.* A pair of iterative Cox regression survival analyses, one with any new serious offense as the dependant variable,

and one with any new offense as the dependant variable, were run with the modified Static-99 entered in the first step and modified SAPROF entered in the second. Cox regression survival analysis provides a means to control for unequal follow-up times in the sample and to appropriately manage censored cases (Wright, 2000). If the modified SAPROF accounts for a significant portion of the variance in recidivism outcomes, and its addition to the model results in a significant increase in the total variance in recidivism outcomes explained by the model, it will indicate that the modified SAPROF has incremental validity over the modified Static-99 (Langton & Worling, 2015; Wright, 2000).

## **Results**

The analyses described below are based on a modified version of the Static-99 which omits items 1 (age at release from index sex offense) and 2 (ever lived with a lover for at least two years) because, as previously noted, they may be developmentally inappropriate for adolescent samples. Similarly, a modified version of the SAPROF total score is used, with items 6 (work), 8 (financial management), and 12 (medication) omitted.

### **Inter-rater Reliability**

Inter-class correlation coefficients (ICCs) were calculated as an index of inter-rater reliability, based on a single-rating with three raters, consistency-agreement, 2-way random-effects model. Cicchetti (1994) suggested that ICCs below .40 constitute poor inter-rater reliability, .40 to .59 constitutes fair inter-rater reliability, .60 to .74 constitutes good interrater reliability, and .75 and above constitutes excellent inter-rater reliability. The ICC value for the modified Static-99 was .84 falling in the excellent range. The modified SAPROF total score, the ICC value was .772, also falling in the excellent range. The SAPROF internal, modified

motivational, and external subscales fell in the good ( $ICC = .61$ ), excellent ( $ICC = .76$ ), and good ( $ICC = .74$ ) range, respectively. See Table 6 for descriptive statistics and ICCs.

**Table 6.**

*Descriptive Statistics and Inter-rater Reliability*

|                              | <i>M</i> | <i>SD</i> | <i>ICC</i> | <i>95% CI</i> |
|------------------------------|----------|-----------|------------|---------------|
| Static-99                    | 1.59     | 1.21      | .84        | .70 - .92     |
| SAPROF Total Score           | 10.73    | 3.67      | .77        | .60 - .89     |
| SAPROF Internal Subscale     | 3.05     | 3.05      | .61        | .38 - .79     |
| SAPROF Motivational Subscale | 3.32     | 1.89      | .76        | .59 - .88     |
| SAPROF External Subscale     | 4.35     | 1.67      | .74        | .55 - .87     |

Note: Static-99 refers to the modified Static-99. SAPROF total score and motivational subscale refer to the modified SAPROF total score and modified SAPROF motivational subscale.

### **Predictive Validity**

The results from the following set of ROC analyses are given in Table 7 (serious recidivism) and 8 (any recidivism). The first hypothesis, that the modified Static-99 would significantly predict both recidivism outcomes (serious and any new offense), was tested using an ROC analysis. Consistent with this hypothesis, the modified Static-99 significantly predicted serious ( $AUC = .63, p = .002, 95\% CI = .55-.71$ ); however, it did not significantly predict any recidivism ( $AUC = .57, p = .061, 95\% CI = .50-.65$ ). Based on Rice and Harris' (2005) descriptions of AUC effect sizes, these effects can be described as small for both serious and any recidivism.

The second hypothesis, that the modified SAPROF total score would significantly predict an absence of serious or any recidivism, was tested using a second pair of ROC analyses.

Consistent with this hypothesis, the modified SAPROF total score significantly predicted an absence of both serious (AUC = .60,  $p = .022$ , 95% CI = .52-.67) and any recidivism (AUC = .61,  $p = .005$ , 95% CI = .53-.68). These effect sizes both fall into the small range (Rice & Harris, 2005).

It was hypothesized that the internal conceptual subscale would significantly predict the absence of both serious and any new offenses during the follow up period. This was tested with a pair of ROC analyses, which indicated that the modified SAPROF internal conceptual subscale did indeed significantly predict an absence of both serious (AUC = .60,  $p = .020$ , 95% CI = .52-.67) and any recidivism (AUC = .63,  $p = .001$ , 95% CI = .56-.70). These effect sizes both fall into the small range (Rice & Harris, 2005).

No hypotheses were put forward for the SAPROF modified motivational and external subscales, given that they have not been found to significantly predict desistance (Klein et al., 2015). ROC analyses were used to determine whether these subscales significantly predict an absence of serious or any recidivism in the present sample. The modified motivational conceptual subscale was found to significantly predict an absence of both serious (AUC = .63,  $p = .002$ , 95% CI = .56-.70) and any (AUC = .61,  $p = .002$ , 95% CI = .55-.68) recidivism, with effect sizes falling into the small range for both outcomes. The external conceptual subscale did not significantly predict an absence of either serious (AUC = .48,  $p = .557$ , 95% CI = .39-.56) or any recidivism (AUC = .49,  $p = .800$ , 95% CI = .42-.57).

Given that the SAPROF external conceptual subscale did not significantly predict an absence of either recidivism outcome, a pair of post-hoc ROC analyses were conducted in order to determine whether removing the external subscale items from the modified SAPROF total score alter the predictive validity of the total score. This was found to be the case, with the

modified SAPROF total score, excluding the external subscale items, significantly predicting the absence of both serious (AUC = .63,  $p = .001$ , 95% CI = .56-.71) and any recidivism (AUC = .64,  $p < .001$ , 95% CI = .57-.71). Notably, these effect sizes fall into the moderate range.

**Table 7.**

*Prediction of Serious Recidivism or Desistance*

| <i>Serious Recidivism</i>                                    |            |           |               |
|--------------------------------------------------------------|------------|-----------|---------------|
|                                                              | <i>AUC</i> | <i>SE</i> | <i>95% CI</i> |
| Static-99                                                    | .63**      | .04       | .55 - .71     |
| <i>Absence of Serious Recidivism</i>                         |            |           |               |
|                                                              | <i>AUC</i> | <i>SE</i> | <i>95% CI</i> |
| SAPROF Total Score                                           | .60*       | .04       | .52 - .67     |
| SAPROF Total Score<br>(excluding external subscale<br>items) | .63**      | .04       | .56 - .71     |
| SAPROF Internal Subscale                                     | .60*       | .04       | .52 - .67     |
| SAPROF Motivational<br>Subscale                              | .63**      | .04       | .56 - .70     |
| SAPROF External Subscale                                     | .48        | .04       | .39 - .56     |

Note: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Static-99 refers to the modified Static-99.

SAPROF total score and motivational subscale refer to the modified SAPROF total score and modified SAPROF motivational subscale.

**Table 8.***Prediction of Any Recidivism or Desistance*

| <i>Any Recidivism</i>                                        |            |           |               |
|--------------------------------------------------------------|------------|-----------|---------------|
|                                                              | <i>AUC</i> | <i>SE</i> | <i>95% CI</i> |
| Static-99                                                    | .57        | .04       | .50 - .65     |
| <i>Absence of Any Recidivism</i>                             |            |           |               |
|                                                              | <i>AUC</i> | <i>SE</i> | <i>95% CI</i> |
| SAPROF Total Score                                           | .61**      | .04       | .53 - .68     |
| SAPROF Total Score<br>(excluding external subscale<br>items) | .64***     | .04       | .57 - .71     |
| SAPROF Internal Subscale                                     | .63**      | .04       | .56 - .70     |
| SAPROF Motivational<br>Subscale                              | .61**      | .04       | .55 - .68     |
| SAPROF External Subscale                                     | .49        | .04       | .42 - .57     |

Note: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Static-99 refers to the modified Static-99.

SAPROF total score and motivational subscale refer to the modified SAPROF total score and modified SAPROF motivational subscale.

**Incremental Validity**

A set of Cox regression analyses were conducted in order to assess the fourth hypothesis: the modified SAPROF total score will have incremental validity over the modified Static-99. First, extended Cox regression models with each predictor and its respective interaction term with time were used to test the proportional hazards assumption. No statistically significant interaction with time was found for any of the predictors for either outcome.

For each recidivism outcome, three models were tested. Given that the SAPROF external subscale was found not to be a significant predictor of an absence of either recidivism outcome, the modified SAPROF total score, excluding the external subscale was used and the incremental validity of the external subscale was not tested. For the set of three models to predict new serious offenses, the modified Static-99 was entered in the first step with the modified SAPROF total score (model 1), SAPROF internal subscale (model 2), and the modified SAPROF motivational

subscale (model 3). The results from these analyses are available in Table 9. For the set of four models to predict any new offenses, the modified Static-99 was again entered in the first step, with the modified SAPROF total score (model 4), SAPROF internal subscale (model 5), and the modified SAPROF motivational subscale (model 6). The statistics for this set of analyses can be viewed in Table 10.

### ***Model 1***

Incremental validity of the modified SAPROF total score over the modified Static-99 was found for serious recidivism. This finding provides support for the fourth hypothesis. In step one, the overall model was significant ( $\chi^2 = 12.77, 1df, p < .001$ ), with the modified Static-99 significantly predicting serious recidivism ( $b = .28, SE = .08, Wald (1) = 13.21, p < .001$ ). The proportional hazards ratio of  $e^b = 1.32$  (95% CI = 1.14, 1.54) suggests that for each one unit increase in modified Static-99 total score, an adolescent in the present sample was 1.32 times more likely to seriously reoffend. In the second step, when the modified SAPROF total score was entered, the overall model continued to be significant ( $\chi^2 = 17.79, 2df, p < .001$ ). The modified Static-99 continued to be significant predictor ( $b = .26, SE = .08, Wald (1) = 11.00, p = .001, e^b = 1.29$  (95% CI = 1.11, 1.50)). The addition of the modified SAPROF total score was significant ( $\Delta\chi^2 = 5.15, 1df, p = .023; b = -0.08, SE = .04, Wald (1) = 5.05, p = .025$ ). The proportional hazards ratio of  $e^b = .92$  (95% CI = .86, .99) indicated that for each increase of one unit in the modified SAPROF total score, an adolescent in this sample was 1.08 times less likely to commit a new serious offense on any given day controlling for risk (1.00/0.92).

### ***Model 2***

Incremental validity of the SAPROF internal conceptual subscale over the modified Static-99 was found for serious recidivism. In the second step, the overall model continued to be

significant ( $\chi^2 = 17.40, 2df, p < .001$ ) and the modified Static-99 continued to be a significant predictor ( $b = .23, SE = .08, Wald (1) = 8.51, p = .004, e^b = 1.26$  (95% CI = 1.08, 1.47)). The addition of the SAPROF internal subscale was significant ( $\Delta\chi^2 = 4.88, 1df, p = .027; b = -0.18, SE = .08, Wald (1) = 4.65, p = .031$ ). The proportional hazards ratio of  $e^b = .84$  (95% CI = .71, .98) suggests that for each increase of one unit in the SAPROF internal subscale score, an adolescent in this sample was 1.19 times less likely to commit a new serious offense on any given day, controlling for risk (1.00/0.84).

### ***Model 3***

The modified SAPROF motivational conceptual subscale was also found to have incremental validity over the modified Static-99 for serious recidivism. After the modified SAPROF motivational conceptual subscale was entered in the second step, the overall model continued to be significant ( $\chi^2 = 20.68, 2df, p < .001$ ), and the modified Static-99 remained a significant predictor ( $b = .23, SE = .08, Wald (1) = 9.40, p = .002, e^b = 1.26$  (95% CI = 1.09, 1.47)). The addition of the modified SAPROF motivational subscale was significant ( $\Delta\chi^2 = 8.02, 1df, p = .005; b = -0.20, SE = .07, Wald (1) = 7.75, p = .005$ ). Further, the proportional hazards ratio ( $e^b = .82$  (95% CI = .71, .94)) indicates that for each one unit increase in a participant's modified SAPROF motivational subscale score, they were 1.22 times less likely to seriously reoffend on any given day, controlling for risk (1.00/0.82).

### ***Model 4***

The modified SAPROF total score was found to have incremental validity over the modified Static-99 for any recidivism, consistent with the fourth hypothesis. As with previous models, the overall model was significant in step one ( $\chi^2 = 7.47, 1df, p = .006$ ) and the modified Static-99 was a significant predictor ( $b = .21, SE = .08, Wald (1) = 7.63, p = .006$ ). The

proportional hazards ratio of  $e^b = 1.24$  (95% CI = 1.06, 1.44) suggests that for each one unit increase in modified Static-99 total score, an adolescent in the present sample was 1.24 times more likely to commit any new offense. In the second step, the modified SAPROF total score was added, and the overall model remained significant ( $\chi^2 = 15.73$ , 2df,  $p < .001$ ). The modified Static-99 continued to be a significant predictor ( $b = .19$ ,  $SE = .08$ , Wald (1) = 6.12,  $p = .013$ ,  $e^b = 1.21$  (95% CI = 1.04, 1.41)). Further, the addition of the modified SAPROF total score was significant ( $\Delta\chi^2 = 8.49$ , 1df,  $p = .004$ ;  $b = -0.09$ ,  $SE = .03$ , Wald (1) = 8.31,  $p = .004$ ), with a proportional hazards ratio of  $e^b = .92$  (95% CI = .86, .97) indicating that for each one unit increase in a participant's modified SAPROF total score, they were 1.09 times less likely to commit any new offense on any given day, controlling for risk (1.00/0.92).

#### ***Model 5***

The SAPROF internal conceptual subscale was found to have incremental validity over the modified Static-99 for the any new offense outcome. After the SAPROF internal subscale was added to the model in the second step, the overall model continued to be significant ( $\chi^2 = 18.14$ , 2df,  $p < .001$ ); however, modified the Static-99 was no longer a significant predictor ( $b = .15$ ,  $SE = .08$ , Wald (1) = 3.56,  $p = .059$ ,  $e^b = 1.16$  (95% CI = 0.99, 1.36)). The addition of the SAPROF internal subscale was significant ( $\Delta\chi^2 = 11.38$ , 1df,  $p = .001$ ;  $b = -0.23$ ,  $SE = .07$ , Wald (1) = 10.64,  $p = .001$ ), with a proportional hazards ratio of  $e^b = .79$  (95% CI = .69, .91), indicating that for each one unit increase in a participant's SAPROF internal subscale score, they were 1.27 times less likely to commit any reoffense on any given day, controlling for risk (1.00/0.79).

#### ***Model 6***

Incremental validity of the modified SAPROF motivational conceptual subscale over the modified Static-99 was found for the any new offense outcome. When the modified SAPROF motivational subscale was added to the model in the second step, the overall model remained significant ( $\chi^2 = 15.38, 2df, p < .001$ ) and the modified Static-99 continued to be a significant predictor ( $b = .18, SE = .08, Wald (1) = 5.84, p = .016, e^b = 1.20$  (95% CI = 1.04, 1.40)). Further, the addition of the modified SAPROF motivational subscale was significant ( $\Delta\chi^2 = 8.08, 1df, p = .004; b = -0.17, SE = .06, Wald (1) = 7.90, p = .005$ ). The proportional hazards ratio ( $e^b = .85$  (95% CI = .76, .95)) indicates that for each one unit increase in a participant's score on the modified motivational subscale, they were 1.18 times less likely to commit any new offense on any given day, controlling for risk (1.00/0.85).

**Table 9.***Hierarchical Models for Serious Recidivism*

|                                                     | <i>b</i> | <i>SE</i> | <b>Wald</b> | <i>e<sup>b</sup></i> | <b>95% CI</b> | $\chi^2$ | $\Delta\chi^2$ |
|-----------------------------------------------------|----------|-----------|-------------|----------------------|---------------|----------|----------------|
| <i>Step 1: Static-99</i>                            |          |           |             |                      |               | 12.77*** |                |
| Static-99                                           | .28***   | .08       | 13.20       | 1.32                 | 1.14, 1.54    |          |                |
| <i>Model 1, Step 2</i>                              |          |           |             |                      |               | 20.85*** | 8.46**         |
| Static-99                                           | .22**    | .08       | 7.76        | 1.24                 | 1.07, 1.45    |          |                |
| SAPROF Total Score<br>(excluding external<br>items) | -0.13**  | .04       | 8.04        | .88                  | .81, .96      |          |                |
| <i>Model 2, Step 2</i>                              |          |           |             |                      |               | 17.40*** | 4.88*          |
| Static-99                                           | .23**    | .08       | 8.51        | 1.26                 | 1.08, 1.47    |          |                |
| SAPROF Internal<br>Subscale                         | -0.18*   | .08       | 4.65        | .84                  | .71, .98      |          |                |
| <i>Model 3, Step 2</i>                              |          |           |             |                      |               | 20.68*** | 8.02**         |
| Static-99                                           | .23**    | .08       | 9.40        | 1.26                 | 1.09, 1.47    |          |                |
| SAPROF<br>Motivational Subscale                     | -0.20**  | .07       | 7.75        | .82                  | .71, .94      |          |                |

Note: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Static-99 refers to the modified Static-99. SAPROF total score and motivational subscale refer to the modified SAPROF total score and modified SAPROF motivational subscale.

**Table 10.***Hierarchical Models for Any Recidivism*

|                                                     | <i>b</i> | <i>SE</i> | <b>Wald</b> | <i>e<sup>b</sup></i> | <b>95% CI</b> | $\chi^2$ | $\Delta\chi^2$ |
|-----------------------------------------------------|----------|-----------|-------------|----------------------|---------------|----------|----------------|
| <i>Step 1: Static-99</i>                            |          |           |             |                      |               | 7.47**   |                |
| Static-99                                           | .21*     | .08       | 7.63        | 1.24                 | 1.06, 1.44    |          |                |
| <i>Model 4, Step 2</i>                              |          |           |             |                      |               | 19.26*** | 12.44***       |
| Static-99                                           | .16*     | .08       | 4.11        | 1.17                 | 1.01, 1.36    |          |                |
| SAPROF Total<br>Score (excluding<br>external items) | -0.13**  | .04       | 11.83       | .88                  | .81, .95      |          |                |
| <i>Model 5, Step 2</i>                              |          |           |             |                      |               | 18.14*** | 11.38**        |
| Static-99                                           | .15      | .08       | 3.56        | 1.16                 | 0.99, 1.36    |          |                |
| SAPROF Internal<br>Subscale                         | -0.23**  | .07       | 10.64       | .79                  | .69, .91      |          |                |
| <i>Model 6, Step 2</i>                              |          |           |             |                      |               | 15.38*** | 8.08**         |
| Static-99                                           | .18*     | .08       | 5.84        | 1.20                 | 1.04, 1.40    |          |                |
| SAPROF<br>Motivational<br>Subscale                  | -0.17**  | .06       | 7.90        | .85                  | .76, .95      |          |                |

Note: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Static-99 refers to the modified Static-99. SAPROF total score and motivational subscale refer to the modified SAPROF total score and modified SAPROF motivational subscale.

## Discussion

The present study aimed to investigate the predictive validity of the risk-based modified Static-99 and strengths-based modified SAPROF among a sample of adolescent males with sexual offenses for two recidivism outcome measures: serious new offenses (including sexual and violent) and any new offense. It was hypothesized that the modified Static-99 would be a significant predictor of both outcomes. This was partially supported, with the modified Static-99 being a significant predictor of the serious recidivism outcome, but not the any recidivism outcome. Although previous work with adults has found evidence indicating that the Static-99 is a valid predictor of any recidivism (Langton et al., 2007a; Turner et al., 2016), it should be noted that the Static-99 is intended to predict sexual recidivism specifically. Given this, the present finding may be due to the Static-99 not being intended for predicting any recidivism. Alternatively, it may be the case that the Static-99 items are predictive of general recidivism for adults (Langton et al., 2007a; Turner et al., 2016), but not adolescents.

It was hypothesized that the modified SAPROF total score and internal conceptual subscale would be significant predictors of an absence of both serious and any recidivism. These hypotheses were supported. Although no hypotheses were put forth for the modified motivational and external conceptual subscales, their predictive validity was also tested for both outcomes. The modified motivational subscale was a significant predictor of an absence of both serious and any recidivism; however, the external subscale was not a significant predictor of either outcome. These findings are in line with previous work with adults, which has shown that the SAPROF total score, internal and motivational subscales, but not external subscale are significant predictors of general, sexual, and violent recidivism (Coupland & Olver, 2020; Yoon et al., 2018). Given that the external subscale was not a significant predictor of either recidivism

outcome, the predictive validity of a version of the modified SAPROF total score was tested, with the external subscale items removed. Notably, this version of the modified SAPROF was a significant predictor of an absence of both serious and any recidivism, with a moderate effect size, whereas the modified SAPROF total score, including the external subscale had a small effect size. These findings provide insight as to factors which may be associated with desistance among adolescent males who have sexually offended. More specifically, factors such as those included in the internal and motivational subscales may be associated with desistance in this population. Conversely, factors associated with the external subscale may not be associated with desistance from either serious or any offending.

Based on previous findings indicating that the SAPROF has incremental validity over another risk-based measure, the HCR-20 (de Vries Robbé et al., 2011, 2013, 2015b), it was hypothesized that the modified SAPROF total score would have incremental validity over the modified Static-99. Given that the external conceptual subscale was not a significant predictor of either outcome, the modified SAPROF total score excluding it was used, and the incremental validity of the external subscale was not tested. The modified SAPROF total score, internal conceptual subscale, and modified motivational subscale were found to have incremental validity over the modified Static-99 for both serious and any recidivism outcomes. These findings indicate that when the modified SAPROF is included in the model, more of the variance in recidivism outcomes is explained than by the modified Static-99 alone. In other words, for this sample, the inclusion of the modified SAPROF provided information regarding likelihood of reoffending beyond that provided by the modified Static-99. This is particularly relevant given that some have argued against attending to strengths because they are merely the opposite of risk factors (Thornton et al., 2017). For example, Thornton and colleagues (2017) assert that a factor

can be either a risk or protective, depending on the role it plays for a given individual. These findings, with the strengths-based modified SAPROF having incremental validity over the risk-based modified Static-99, provide support for the inclusion of strengths in forensic assessments.

Of note, while these findings indicate that the modified Static-99 and modified SAPROF are valid predictors of serious and any recidivism in the present sample, it cannot be concluded based on these findings that either tool should be used with adolescent samples in clinical practice. The developers of both the Static-99(R) (Hanson & Thornton, 2000; Phenix et al., 2016) and the SAPROF (de Vogel et al., 2011) state that these tools should be used exclusively with adult populations. In addition to the developers' specifications regarding which populations these tools can be used with, clinicians are ethically required to select assessment tools which are suited to the characteristics of the individual being assessed (CPA, 2017, Section II.18). There are several alternative tools which were developed specifically for use with adolescents. For example, the Juvenile Sexual Offense Recidivism Risk Assessment Tool – II (JSORRAT-II; Epperson et al., 2006) is a risk-based tool that has been shown to be a significant predictor of any sexual (AUC = .65) and sexually violent (AUC = .65) recidivism among a sample of adolescents who have sexually offended (Epperson & Ralston, 2015). Additionally, there is a youth version of the SAPROF (SAPROF-YV; de Vries Robbé et al., 2015). There have been few studies investigating the predictive validity of this tool; however, Kleeven and colleagues (2020) found that the SAPROF-YV total score significantly predicted an absence of both general (AUC = .67) and violent (AUC = .65) recidivism after a follow up period of two years, in sample of justice-involved adolescent males.

It should be noted that, as there is overlap in risk factors for recidivism among adult and adolescent populations, and in protective factors too with these two populations (Arnold &

Davis, 2016; Craig et al., 2005; Efta-Breitbach & Freeman, 2004; Hanson & Harris, 2000; Långström, 2002; Thornton, 2013; Worling & Långström, 2003, 2006), items in both risk and strengths-based tools intended for adults and those intended for use with adolescents also overlap. For example, the Static-99 and JSORRAT-II both assess history of non-sexual violence (with the Static-99 “prior non-sexual violence” item and JSORRAT-II “number of adjudications for non-sexual violence” item) and history of sexual offending (with the Static-99 “prior sex offenses” and JSORRAT-II “number of adjudications for sex offenses” items) (Epperson et al., 2006; Hanson & Thornton, 2000). The Static-99 and Juvenile Sex Offender Assessment Protocol – II (J-SOAP-II; Prentky & Righthand, 2003) also assesses prior sexual offenses, as well as victim characteristics (with the Static-99 item “any male victims” and J-SOAP item “male child victim”) (Hanson & Thornton, 2000; Prentky & Righthand, 2003). Of note, these overlapping items assess risk factors associated with recidivism among those who have sexually offended: prior sexual offending and male victims (Arnold & Davis, 2016; Craig et al., 2005; Efta-Breitbach & Freeman, 2004; Hanson & Harris, 2000; Långström, 2002; Thornton, 2013; Worling & Långström, 2003, 2006).

Overlap also exists for strengths-based tools. The SAPROF has a number of overlapping items with the Structured Assessment of Violence Risk in Youth (SAVRY; Borum et al., 2006), assessing attachment (the SAPROF item “secure attachment in childhood” and the SAVRY item “strong attachment and bonds”), attitudes towards treatment and authority (the SAPROF items “motivation for treatment” and “attitudes towards authority” and the SAVRY item “positive attitude towards treatment and authority”), and one’s social network (the SAPROF item “social network” and SAVRY item “social support”) (Borum et al., 2006; de Vogel et al., 2011). The SAPROF and SAPROF-YV have a considerable number of items similar in focus, including:

coping, self-control, motivation for treatment, leisure activities, medication, and professional care (de Vogel et al., 2011; de Vries Robbé et al., 2015). Evidently, there is considerable overlap in factors assessed by both risk and strengths-based tools intended for use with adults and those intended for use with adolescents. This overlap may aid in explaining why measures developed for use with adult populations (specifically the Static-99 and SAPROF) significantly predict outcomes when used with adolescent samples.

Despite the considerable overlap in focus for some of these items, their operationalizations are not all highly similar. Furthermore, items in assessment tools developed for use with adults may not be sufficiently sensitive to developmental considerations that are important when considering adolescents. This is illustrated by the “age at release” item in the Static-99 and Static-99R. Given that all individuals below the age of 18 receive the same score (Hanson & Thornton, 2000; Phenix et al., 2016), this item is not sensitive when used with adolescents, as it is unable to capture differences in risk between age categories below 18 years. Although this item assesses a risk factor (young age) that has been found to be associated with recidivism in both adults and adolescents who have sexually offended (Arnold & Davis, 2016; Craig et al., 2005; Heilbrun et al., 2005; Plattner et al., 2009), the item as operationalized cannot be used with adolescents because anyone below the age of 18 would receive the same score, despite evidence indicating that age is a relevant consideration with adolescents, with recidivism risk found to be higher for younger adolescents (Heilbrun et al., 2005; Plattner et al., 2009).

Although it cannot be concluded from these findings that these tools should be used in clinical forensic practice with adolescents, the results of the present study are quite promising in terms of their implications for the understanding of both risk and protective factors as they relate to recidivism and desistance in adolescents who have offended sexually. As noted above, the

finding that the modified Static-99 is not a significant predictor of any recidivism in the present sample is incongruent with prior research using adult samples, which has found support for the predictive validity of the Static-99 for general recidivism (Langton et al., 2007a; Turner et al., 2016). This suggests that although risk factors such as those included in the Static-99 may be useful for predicting any recidivism in adults (Langton et al., 2007a; Turner et al., 2016), such factors may not be similarly useful with adolescents or at least may need to be operationalized differently in some way perhaps in terms of the weighting/coding of each item. These findings also indicate that adults and adolescents who have sexually offended share risk factors for serious recidivism, as well as protective factors against both serious and any recidivism. Altogether, the present findings expand upon current understanding for both risk and protective factors for any and serious recidivism among adolescents who have sexually offended, and how these factors compare to those for adults.

The findings supporting the predictive and incremental validity of a modified version of the SAPROF when used with adolescents who have sexually offended are particularly noteworthy with regards to their implications for informing intervention. Unlike the Static-99, which assesses static factors that cannot be changed via intervention, the SAPROF is notable because it focuses primarily on dynamic strengths, which can be targeted in treatment. Although, as noted above, it cannot be concluded from these findings that the SAPROF should be used in forensic clinical assessments with adolescents in order to make judgements about an individuals' likelihood of reoffending, the modified SAPROF, specifically the internal and modified motivational subscales, may be promising as sets of items that could inform treatment. The SAPROF items provide information regarding strengths an individual already possesses, and importantly, those which can be targeted in treatment, due to their dynamic nature. Thus, given

that the modified SAPROF total score (excluding the external subscale) was found to significantly predicts both an absence of serious and any recidivism among adolescents who have sexually offended, further research is clearly warranted to determine which of the items are predictive of (types of) recidivism and whether change in any corresponds to a change in the likelihood of (types of) recidivism. Such information would be valuable in intervention planning.

In contrast to the tendency in the literature to focus on risks, positive psychology approaches involve a focus on the individual's strengths and well-being (Parks et al., 2015; Parks & Schueller, 2014). One positive psychology approach, the Good Lives Model (GLM), posits that forensic assessment should include a determination of which primary and secondary goods an individual already possesses, and those they aspire to attain (Ward & Fortune, 2013; Ward & Mann, 2004). Primary goods refer to outcomes which contribute to well-being, such as knowledge and mastery, and secondary goods refer to means by which one can obtain primary goods, such as joining a sports team to obtain mastery and friendship (Ward & Mann, 2004). Some proponents of the GLM advocate not only for the inclusion of strengths in forensic assessment and intervention, but for the sole inclusion of strengths and the exclusion of risks (Heffernan & Ward, 2019; Ward, 2002). For example, while describing the GLM, Ward (2002) explained that human well-being, rather than risk management should be the primary focus of rehabilitation. Although it has been noted that the current findings provide some support for the inclusion of strengths in forensic practice, it should be stated that the findings of the present study do not support an exclusive focus on strengths. Indeed, the present findings indicate that both strengths and risks are significant predictors of recidivism, and the incremental validity of the modified SAPROF over the modified Static-99 suggests that strengths and risks both represent important information that (at least when operationalized and demonstrated to be

dynamic in nature) has real world significance in assessment and intervention work. Given this, exclusive attention to either alone may result in incomplete case conceptualizations.

### **Limitations**

The present study included participants from three assessments sites: a private practice, a Ministry program, and a community agency all within a confined geographical area. It is possible that participants' scores on either assessment tool or recidivism outcomes may have varied significantly between these sites. This was, however, a deliberate decision in planning the construction of the sample in order to ensure an adequate sample size and as wide a range of scores as possible on the tools.

A second potential limitation of the present studies involves the recidivism outcomes used. More specifically, as previously noted, neither the Static-99 nor the SAPROF are intended to predict any recidivism (de Vogel et al., 2011; Phenix et al., 2016). Further, although predictive accuracy and recidivism rates for violent (including sexual) recidivism as well as sexual recidivism was originally reported for the Static-99 (Hanson & Thornton, 2000), its developers are explicit that it should be used now only for assessing risk for sexual recidivism (Phenix et al., 2016). The small effect sizes found for the modified SAPROF predicting the absence of any recidivism and the non-significant prediction of any recidivism found for the modified Static-99 may simply be due to this outcome not being one that the tools were originally developed to predict.

### **Future Directions**

There is a strong tendency in the literature to focus on risks, and a notable dearth of research investigating factors protective against recidivism. Continued attention to protective factors and strengths-based assessment tools in future research would be advantageous for a few

reasons. First, greater attention to protective factors and strengths-based tools in the literature will improve our understanding of these factors and how they contribute to desistance from a variety of types of offending across various populations. Second, a larger body of research investigating the predictive validity of strengths-based tools may provide support for their use in assessment and treatment. More widespread inclusion of strengths in both assessment and treatment has several advantages. As previously noted, strengths-based tools such as the SAPROF and SAPROF-YV are particularly useful for informing treatment as the items they are made up of can be targeted by intervention (de Vogel et al., 2011; de Vries Robbé et al., 2015). Further, the inclusion of strengths in clinical forensic assessment and treatment aligns with positive psychology, which has been demonstrated to be an effective approach to intervention in forensic contexts (Parks et al., 2015; Parks & Schueller, 2014). Finally, previous research has found that protective factors may buffer the impact of risks on one's likelihood of reoffending (Jones et al., 2015, 2016; Lodewijks et al., 2010). This, coupled with findings from the present study and others indicating that certain strengths-based-measures may have incremental validity over certain risk-based measures (de Vries Robbé et al., 2011, 2013, 2015b), suggest that future research investigating the impact of the inclusion of strengths in assessments on their accuracy is warranted.

The present study is one of a small number of which have investigated the incremental validity of strengths-based tools over risk-based ones (de Vries Robbé et al., 2011, 2013, 2015b). Previous work has found support for the incremental validity of the SAPROF over the HCR-20 among a sample of adults who are patients in a forensic psychiatric facility (de Vries Robbé et al., 2011, 2013, 2015b), and the present study established evidence for the incremental validity of a modified version of the SAPROF over a modified version of the Static-99 among a sample

of adolescent males. Given these findings and their value in terms of bolstering our understanding risk and protective factors, discussed above, future research investigating the incremental validity of strengths-based tools intended for use with adolescents, such as the SAPROF-YV, over risk-based measures also intended for this population, such as the JSORRAT-II (Epperson et al., 2006) is warranted. Findings from such an investigation would further our understanding of risk and protective factors among adolescents who have offended and, if strengths-based tools such as the SAPROF-YV are found to have incremental validity over risk-based tools, provide further evidence that strengths can provide non-overlapping information alongside risks. Notably, given that such tools are meant for use with adolescent populations, such findings may inform forensic assessment by providing support for the use of strengths-based instruments in combination with risk-based ones.

### **Conclusion**

The present study investigated the predictive validity of a modified version of two tools intended for use with adults, the risk-based Static-99 and strengths-based SAPROF, among a sample of adolescent males who have sexually offended. Evidence was found for the predictive validity of the modified SAPROF total score and SAPROF internal and modified motivational subscales for both serious and any recidivism. Support was also found for the predictive validity of a modified version of the Static-99 for serious recidivism. The incremental validity of the modified SAPROF over the modified Static-99 was also investigated, and evidence for this was found for both serious and any recidivism. These findings bolster our understanding of risk and protective factors for serious and any recidivism among adolescents who have sexually offended, as well as how these factors vary between adolescents and adults. Finally, the present findings

have implications for forensic clinical practice, particularly with regards to the use of SAPROF items to inform intervention.

## References

- Abidin, Z., Davoren, M., Naughton, L., Gibbons, O., Nulty, A., & Kennedy, H. G. (2013). Susceptibility (risk and protective) factors for in-patient violence and self-harm: prospective study of structured professional judgement instruments START and SAPROF, DUNDRUM-3 and DUNDRUM-4 in forensic mental health services. *BMC psychiatry*, 13(1), 1-18. <https://doi.org/10.1186/1471-244X-13-197>
- Allen, M. K., & Superle, T. (2016). Youth crime in Canada, 2014. *Juristat: Canadian Centre for Justice Statistics*, 1-50. <https://www150.statcan.gc.ca/n1/pub/85-002-x/2016001/article/14309-eng.pdf>
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *The American Psychologist*, 68(1), 7-19. <https://www.apa.org/pubs/journals/features/forensicpsychology.pdf>
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2000). *Level of service/case management inventory: LS/CMI*. Toronto, Canada: Multi-Health Systems.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of personality assessment*, 87(1), 84-94. [https://doi.org/10.1207/s15327752jpa8701\\_07](https://doi.org/10.1207/s15327752jpa8701_07)
- Arnold, D., & Davis, M. (2016). Risk Factors and Risk Assessments for Sexual Offense Recidivism. In *Sexual Offending* (pp. 417-435). Springer, New York, NY. [https://doi.org/10.1007/978-1-4939-2416-5\\_18](https://doi.org/10.1007/978-1-4939-2416-5_18)
- Austin, A. E., & Short, N. A. (2020). Sexual violence, mental health, and prescription opioid use and misuse. *American Journal of Preventative Medicine*, 59(6), 818-827. <https://doi.org/10.1016/j.amepre.2020.06.019>

- Babchishin, K. M., Blais, J., & Helmus, L. (2012). Do static risk factors predict differently for Aboriginal sex offenders? A multi-site comparison using the original and revised Static 99 and Static-2002 scales. *Canadian Journal of Criminology and Criminal Justice*, 54(1), 1-43. <https://doi.org/10.3138/cjccj.2010.E.40>
- Barbaree, H. E., Langton, C. M., Blanchard, R., & Cantor, J. M. (2009). Aging versus stable enduring traits as explanatory constructs in sex offender recidivism: Partitioning actuarial prediction into conceptually meaningful components. *Criminal Justice and Behavior*, 36(5), 443-465. <https://doi.org/10.1177/0093854809332283>
- Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal justice and behavior*, 28(4), 490-521. <https://doi.org/10.1177/009385480102800406>
- Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology*, 47(4), 422-438. <https://doi.org/10.1177/0306624X03253850>
- Beauregard, E., & Mieczkowski, T. (2009). Testing the predictive utility of the STATIC-99: A Bayes analysis. *Legal and Criminological Psychology*, 14(2), 187-200. <https://doi.org/10.1348/135532508X330994>
- Bengtson, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse*, 19(2), 135-153. <https://doi.org/10.1177/107906320701900205>
- Boccaccini, M. T., Rice, A. K., Helmus, L. M., Murrie, D. C., & Harris, P. B. (2017). Field validity of Static-99/R scores in a statewide sample of 34,687 convicted sexual

offenders. *Psychological Assessment*, 29(6), 611-623.

<https://doi.org/10.1037/pas0000377>

Boer, D. P., Wilson, R. J., Gauthier, C. M., & Hart, S. D. (1997). Assessing risk of sexual violence: Guidelines for clinical practice. *Impulsivity: Theory, assessment, and treatment*, 326-342.

Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)*. Odessa: Psychological Assessment Resources.

Bushway, S. D., & Paternoster, R. (2013). Desistance from crime: A review and ideas for moving forward. In *Handbook of life-course criminology* (pp. 213-231). Springer.

[https://doi.org/10.1007/978-1-4614-5113-6\\_13](https://doi.org/10.1007/978-1-4614-5113-6_13)

Canadian Psychological Association. (2017a). Canadian Psychological Association Code of Ethics (4<sup>th</sup> ed.). <http://www.cpa.ca/aboutcpa/committees/ethics/codeofethics/>

Chu, C. M., Xu, X., Li, D., Ruby, K., & Chng, G. S. (2020). The Utility of SAPROF-YV Ratings for Predicting Recidivism in Male Youth Under Community Supervision in Singapore. *Criminal Justice and Behavior*, 47(11), 1409-1427.

<https://doi.org/10.1177/0093854820949595>

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, 6(4), 284.

<https://doi.org/10.1037/1040-3590.6.4.284>

Coid, J. W., Kallis, C., Doyle, M., Shaw, J., & Ullrich, S. (2015). Identifying causal risk factors for violence among discharged patients. *PloS one*, 10(11), e0142493.

<https://doi.org/10.1371/journal.pone.0142493>

Coupland, R., & Olver, M. E. (2020). Assessing protective factors in treated violent offenders:

- Associations with recidivism reduction and positive community outcomes. *Psychological assessment*, 32(5), 493-508. <http://doi.org/10.1037/pas0000807>
- Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression*, 11(1), 65-84. <https://doi.org/10.1080/13552600410001667733>
- de Vogel, V., Bruggeman, M., & Lancel, M. (2019). Gender-sensitive violence risk assessment: predictive validity of six tools in female forensic psychiatric patients. *Criminal justice and behavior*, 46(4), 528-549. <https://doi.org/10.1177/0093854818824135>
- de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y. H. (2011). Assessing protective factors in forensic psychiatric practice: Introducing the SAPROF. *International journal of forensic mental health*, 10(3), 171-177. <https://doi.org/10.1080/14999013.2011.600230>
- de Vries Robbé, M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *Journal of Forensic Psychiatry and Psychology*, 24(4), 440-457. <https://doi.org/10.1080/14789949.2013.818162>
- de Vries Robbé, M., de Vogel, V., & de Spa, E. (2011). Protective factors for violence risk in forensic psychiatric patients: A retrospective validation study of the SAPROF. *International journal of forensic mental health*, 10(3), 178-186. <https://doi.org/10.1080/14999013.2011.600232>
- de Vries Robbé, M., de Vogel, V., Wever, E. C., Douglas, K. S., & Nijman, H. L. I. (2016). Risk and protective factors for inpatient aggression. *Criminal Justice and Behavior*, 43(10), 1364-1385. <https://doi.org/10.1177/0093854816637889>
- de Vries Robbé, M., de Vogel, V., Koster, K., & Bogaerts, S. (2015b). Assessing protective

factors for sexually violent offending with the SAPROF. *Sexual Abuse*, 27(1), 51-70.

<https://doi.org/10.1177/1079063214550168>

- de Vries Robbé, M., Geers, M. C. K., Stapel, M., Hilterman, E. L. B., & de Vogel, V. (2015a). *SAPROF–Youth version: Guidelines for the assessment of protective factors for violence risk in juveniles. English version*. Utrecht: De Forensische Zorgspecialisten.
- Douglas, K. S. (2019). Evaluating and managing risk for violence using structured professional judgment. In D. L. Polaschek, A. Day, C. R. Hollin (Eds.), *The Wiley International Handbook of Correctional Psychology* (pp. 427-445). John Wiley & Sons.
- <https://doi.org/10.1002/9781119139980.ch26>
- Douglas, K. S., Ogloff, J. R., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: the HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of consulting and clinical psychology*, 67(6), 917-930. <https://doi.org/10.1037/0022-006X.67.6.917>
- Efta-Breitbach, J., & Freeman, K. A. (2004). Recidivism and resilience in juvenile sexual offenders: An analysis of the literature. *Journal of Child Sexual Abuse*, 13(3-4), 257-279.
- [https://doi.org/10.1300/J070v13n03\\_13](https://doi.org/10.1300/J070v13n03_13)
- Epperson, D. L., Ralston, C. A., Fowers, D., DeWitt, J., & Gore, K. S. (2006). Actuarial risk assessment with juveniles who offend sexually: Development of the Juvenile Sexual Offense Recidivism Risk Assessment Tool–II. In D. Prescott (Ed.), *Risk assessment of youth who have sexually abused: Theory, controversy, and emerging strategies* (pp. 118-169). Wood & Barnes.
- Epperson, D. L., & Ralston, C. A. (2015). Development and validation of the Juvenile Sexual Offense Recidivism Risk Assessment Tool–II. *Sexual Abuse*, 27(6), 529-558.

<https://doi.org/10.1177/1079063213514452>

- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. John Wiley and Sons. *Acta Physiol Scand*. <https://doi.org/10.1002/bimj.4710300308>
- Grubin, D. (1998). *Sex offending against children: Understanding the risk* (Vol. 99). London, England: Home Office, Policing and Reducing Crime Unit, Research, Development and Statistics Directorate.
- Guyton, M. R., & Jackson, R. L. (2008). Violence risk assessment. In R. Jackson (Ed.), *International perspectives on mental health. Learning forensic assessment* (pp. 153–181). Routledge/Taylor & Francis Group.
- Hanson, R. K. (1990). The psychological impact of sexual assault on women and children: A review. *Annals of Sex Research*, 3(2), 187-232.  
<https://doi.org/10.1177/107906329000300204>
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (No. 1997). Solicitor General Canada.
- Hanson, R. K. (2002). Recidivism and age: Follow-up data from 4,673 sexual offenders. *Journal of interpersonal violence*, 17(10), 1046-1062. <https://doi.org/10.1177/08862605-0201710-02>
- Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of consulting and clinical psychology*, 66(2), 348-362.  
<https://doi.org/10.1037/0022-006X.66.2.348>
- Hanson, R. K., & Harris, A. J. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and behavior*, 27(1), 6-35.  
<https://doi.org/10.1177/0093854800027001002>

- Hanson, R. K., Harris, A. J., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Vol. 5, No. 6). Ottawa, Ontario: Public Safety Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological assessment, 21*(1), 1-21. <https://doi.org/10.1037/a0014421>
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human behavior, 24*(1), 119-136. <https://doi.org/10.1023/A:1005482921333>
- Harris, A. J. R., & Hanson, R. K. (2010). Clinical, actuarial and dynamic risk assessment of sexual offenders: Why do things keep changing?. *Journal of Sexual Aggression, 16*(3), 296-310. <https://doi.org/10.1080/13552600.2010.494772>
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal justice and behavior, 20*(4), 315-335. <https://doi.org/10.1177/0093854893020004001>
- Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). The risk for sexual violence protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence. Vancouver, British Columbia: Mental Health, Law, and Policy Institute, Simon Fraser University and British Columbia Institute on Family Violence.
- Heffernan, R., & Ward, T. (2017). A comprehensive theory of dynamic risk and protective factors. *Aggression and violent behavior, 37*, 129-141. <https://doi.org/10.1016/j.avb.2017.10.003>
- Heilbrun, K., Lee, R., & Cottle, C. C. (2005). Risk factors and intervention outcomes. In K.

- Heilbrun, N. E., Sevin Goldstein, & R. E. Redding (Eds.), *Juvenile delinquency: Prevention, assessment, and intervention* (pp. 111-133).
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse, 24*(1), 64-101. <https://doi.org/10.1177/1079063211409951>
- Hoddenbagh, J., McDonald, S. E., & Zhang, T. (2014). *An estimation of the economic impact of violent victimization in Canada, 2009*. Research and Statistics Division, Department of Justice Canada. [https://www.justice.gc.ca/eng/rp-pr/cj-jp/victim/rr14\\_01/index.html](https://www.justice.gc.ca/eng/rp-pr/cj-jp/victim/rr14_01/index.html)
- Hoge, R. D. (2005). *Youth Level of Service/Case Management Inventory*. In T. Grisso, G. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 283–294). The Guilford Press.
- Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2015). Incorporating strengths into quantitative assessments of criminal risk for adult offenders: The Service Planning Instrument. *Criminal Justice and Behavior, 42*(3), 321-338. <https://doi.org/10.1177/0093854814547041>
- Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2016). Validity of the youth assessment and screening instrument: A juvenile justice tool incorporating risks, needs, and strengths. *Law and human behavior, 40*(2), 182. <https://doi.org/10.1037/lhb0000170>
- Kashiwagi, H., Kikuchi, A., Koyama, M., Saito, D., & Hirabayashi, N. (2018). Strength-based assessment for future violence risk: a retrospective validation study of the Structured Assessment of PROtective Factors for violence risk (SAPROF) Japanese version in forensic psychiatric inpatients. *Annals of general psychiatry, 17*(1), 1-8. <https://doi.org/10.1186/s12991-018-0175-5>

- Kleeven, A. T., De Vries Robbé, M., Mulder, E. A., & Popma, A. (2020). Risk assessment in juvenile and young adult offenders: Predictive validity of the SAVRY and SAPROF YV. *Assessment, 0*(0), 1-17. <https://doi.org/10.1177/1073191120959740>
- Klein, V., Rettenberger, M., Yoon, D., Köhler, N., & Briken, P. (2015). Protective factors and recidivism in accused juveniles who sexually offended. *Sexual Abuse, 27*(1), 71-90. <https://doi.org/10.1177/1079063214554958>
- Kroner, D. G. (2019). The Roles of the Risk Estimate and Clinical Information in Risk Assessments. In D. L. Polaschek, A. Day, C. R. Hollin (Eds.), *The Wiley International Handbook of Correctional Psychology* (pp. 446-460). John Wiley & Sons. <https://doi.org/10.1002/9781119139980.ch27>
- Långström, N. (2002). Long-term follow-up of criminal recidivism in young sex offenders: Temporal patterns and risk factors. *Psychology, Crime and Law, 8*(1), 41-58. <https://doi.org/10.1080/10683160208401808>
- Langton, C. M., Barbaree, H. E., Hansen, K. T., Harkins, L., & Peacock, E. J. (2007b). Reliability and validity of the Static-2002 among adult sexual offenders with reference to treatment status. *Criminal Justice and Behavior, 34*(5), 616-640. <https://doi.org/10.1177/0093854806296851>
- Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007a). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior, 34*(1), 37-59. <https://doi.org/10.1177/0093854806291157>
- Langton, C. M., & Worling, J. R. (2015). Introduction to the special issue on factors positively

- associated with desistance for adolescents and adults who have sexually offended. *Sexual Abuse*, 27(1), 3-15. <https://doi.org/10.1177/1079063214568423>
- Langton, C. M., Worling, J. R., & Sheinin, G. (2021). *Predicting recidivism versus desistance among adolescent males with sexual offenses using indices of risk and strength: Main effects and incremental validity*. Manuscript submitted for publication.
- Lee, S. C., Hanson, R. K., & Blais, J. (2020). Predictive accuracy of the Static-99R and Static 2002R risk tools for identifying Indigenous and White individuals at high risk for sexual recidivism in Canada. *Canadian Psychology/Psychologie Canadienne*, 61(1), 42-57. <https://doi.org/10.1037/cap0000182>
- Lodewijks, H. P., de Ruiter, C., & Doreleijers, T. A. (2010). The impact of protective factors in desistance from violent reoffending: A study in three samples of adolescent offenders. *Journal of interpersonal violence*, 25(3), 568-587. <https://doi.org/10.1177/0886260509334403>
- Logan, C. (2016). Structured professional judgment: Applications to sexual offender risk assessment and management. In A. Phenix & H. Hoberman (Eds.), *Sexual offending* (pp. 571-588). Springer, New York, NY. [https://doi.org/10.1007/978-1-4939-2416-5\\_26](https://doi.org/10.1007/978-1-4939-2416-5_26)
- Looman, J., Morphett, N. A., & Abracen, J. (2013). Does consideration of psychopathy and sexual deviance add to the predictive validity of the static-99R?. *International journal of offender therapy and comparative criminology*, 57(8), 939-965. <https://doi.org/10.1177/0306624X12444839>
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of consulting and clinical psychology*, 62(4), 783-792. <https://doi.org/10.1037/0022-006X.62.4.783>

- Moreau, G. (2019). Police-reported crime statistics in Canada, 2018. *Juristat: Canadian Centre for Justice Statistics*, 1-67. [https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2020001/article/00010-eng.pdf?st=yL\\_3LUuK](https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2020001/article/00010-eng.pdf?st=yL_3LUuK)
- Mori, T., Takahashi, M., & Kroner, D. G. (2017). Can unstructured clinical risk judgment have incremental validity in the prediction of recidivism in a non-Western juvenile context?. *Psychological services*, 14(1), 77-86. <https://doi.org/10.1037/ser0000107>
- Orbis Partners. (2000). Youth Assessment Screening Inventory (YASI). *Ottawa, Ontario, Canada: Author.*
- Parks, A. C., Kleiman, E. M., Kashdan, T. B., Hausmann, L. R. M., Meyer, P. S., Day, A. M., Spillane, N. S., & Kahler, C. W. (2015). *Positive psychotherapeutic and behavioral interventions*. In D. V. Jeste & B. W. Palmer (Eds.), *Positive psychiatry: A clinical handbook* (pp. 147–165). American Psychiatric Publishing, Inc. <https://doi.org/10.1176/appi.books.9781615370818.dj08>
- Parks, A. C., & Schueller, S. (Eds.). (2014). *The Wiley Blackwell handbook of positive psychological interventions*. John Wiley & Sons. <https://doi.org/10.1002/9781118315927>
- Passmore, J., & Oades, L. G. (2015). Positive psychology techniques: positive case conceptualisation. *The Coaching Psychologist*, 11(1), 43-45.
- Paternoster, R., & Bachman, R. (2017). Human agency: The missing element in theories of desistance. In A. Blockand & V. van der Geest (Eds.), *The Routledge International Handbook of Life-Course Criminology* (pp. 29-42). Routledge.
- Persson, M., Belfrage, H., Fredriksson, B., & Kristiansson, M. (2017). Violence during imprisonment, forensic psychiatric care, and probation: Correlations and predictive validity of the risk assessment instruments COVR, LSI-R, HCR-20V3, and

SAPROF. *International journal of forensic mental health*, 16(2), 117-129.

<https://doi.org/10.1080/14999013.2016.1266420>

Phenix, A., & Epperson, D. L. (2016). Overview of the development, reliability, validity, scoring, and uses of the Static-99, Static-99R, Static-2002, and Static-2002R. In A. Phenix & H. Hoberman (Eds.), *Sexual Offending* (pp. 437-455). Springer.

[https://doi.org/10.1007/978-1-4939-2416-5\\_19](https://doi.org/10.1007/978-1-4939-2416-5_19)

Phenix, A., Fernandex, Y., Harris, A. J. R., Helmus, M., Hanson, R. K., Thornton, D. (2016). Static-99-R Coding Rules.

[http://www.static99.org/pdfdocs/Coding\\_manual\\_2016\\_v2.pdf](http://www.static99.org/pdfdocs/Coding_manual_2016_v2.pdf)

Plattner, B., Steiner, H., The, S. S., Kraemer, H. C., Bauer, S. M., Kindler, J., Friedrich, M. H., Kasper, S., & Feucht, M. (2009). Sex-specific predictors of criminal recidivism in a representative sample of incarcerated youth. *Comprehensive Psychiatry*, 50(5), 400-407.

<https://doi.org/10.1016/j.comppsy.2008.09.014>

Poole, D., Liedecke, D., & Marbibi, M. (2000). Risk assessment and recidivism in juvenile sexual offenders: A validation study of the Static-99. *Austin, TX: Texas Youth Commission monograph*.

Prentky, R., & Righthand, S. (2003). Juvenile Sex Offender Assessment Protocol-II (J-SOAP-II). *Manual. NCJ, 202316*.

Quesada, S. P., Calkins, C., & Jeglic, E. L. (2014). An examination of the interrater reliability between practitioners and researchers on the Static-99. *International Journal of Offender Therapy and Comparative Criminology*, 58(11), 1364-1375.

<https://doi.org/10.1177/0306624X13495504>

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders:*

- Appraising and managing risk*. American Psychological Association. <https://doi.org/10.1037/11367-000>
- Ralston, C. A., & Epperson, D. L. (2013). Predictive validity of adult risk assessment tools with juveniles who offended sexually. *Psychological assessment*, 25(3), 905-916. <https://doi.org/10.1037/a0032683>
- Ralston, C. A., Epperson, D. L., & Edwards, S. R. (2016). Cross-validation of the JSORRAT-II in Iowa. *Sexual Abuse*, 28(6), 534-554. <https://doi.org/10.1177/1079063214548074>
- Rettenberger, M., & Hucker, S. J. (2011). Structured professional guidelines: International applications. In D. Boer, R. Eher, L. Craig, M. Miner & F. Pfafflin (Eds.), *International perspectives on the assessment and treatment of sexual offenders: Theory, practice, and research* (pp. 85-110). John Wiley & Sons. <https://doi.org/10.1002/9781119990420>
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of consulting and clinical psychology*, 63(5), 737-748. <https://doi.org/10.1037/0022-006X.63.5.737>
- Rich, P. (2009). *Juvenile sexual offenders: A comprehensive guide to risk evaluation*. John Wiley & Sons.
- Shapland, J., & Bottoms, A. (2017). Offending and offence patterns in the early stages of desistance. In A. Blokland & V. van der Geest (Eds.), *The Routledge International Handbook of Life-Course Criminology* (pp. 301-323). Routledge.
- Singh, J. P. (2012). The history, development, and testing of forensic risk assessment tools. In E. Grigorenko (Eds.), *Handbook of juvenile forensic psychology and psychiatry* (pp. 215-225). Springer. [https://doi.org/10.1007/978-1-4614-0905-2\\_14](https://doi.org/10.1007/978-1-4614-0905-2_14)

- Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical psychology review, 31*(3), 499-513.  
<https://doi.org/10.1016/j.cpr.2010.11.009>
- Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry & Psychology, 16*(1), 92-108. <https://doi.org/10.1080/1478994042000270247>
- Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the Static-99 in practice. *Sexual Abuse, 24*(3), 289-302.  
<https://doi.org/10.1177/1079063211423943>
- Swinburne Romine, R. E., Miner, M. H., Poulin, D., Dwyer, S. M., & Berg, D. (2012). Predicting reoffense for community-based sexual offenders: An analysis of 30 years of data. *Sexual Abuse, 24*(5), 501-514. <https://doi.org/10.1177/1079063212446514>
- Thornton, D. (2013). Implications of our developing understanding of risk and protective factors in the treatment of adult male sexual offenders. *International journal of behavioral consultation and therapy, 8*(3-4), 62-65. <http://doi.org/10.1037/h0100985>
- Thornton, D., Kelley, S. M., & Nelligan, K. E. (2017). Protective factors and mental illness in men with a history of sexual offending. *Aggression and violent behavior, 32*, 29-36.  
<https://doi.org/10.1016/j.avb.2016.12.003>
- Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review, 33*(2), 287-316.

<https://doi.org/10.1016/j.cpr.2012.12.002>

Turner, D., Rettenberger, M., Yoon, D., Klein, V., Eher, R., & Briken, P. (2016). Risk assessment in child sexual abusers working with children. *Sexual Abuse, 28*(6), 572-596.

<https://doi.org/10.1177/1079063214564390>

Vandemark, L. M., & Mueller, M. (2008). Mental health after sexual violence: The role of behavioral and demographic risk factors. *Nursing Research, 57*(3), 175-181.

<https://doi.org/10.1097/01.NNR.0000319498.44499.53>

Varela, J. G., Boccaccini, M. T., Murrie, D. C., Caperton, J. D., & Gonzalez Jr, E. (2013). Do the Static-99 and Static-99R perform similarly for White, Black, and Latino sexual offenders?. *International Journal of Forensic Mental Health, 12*(4), 231-243.

<https://doi.org/10.1080/14999013.2013.846950>

Viljoen, J. L., Elkovitch, N., Scalora, M. J., & Ullman, D. (2009). Assessment of reoffense risk in adolescents who have committed sexual offenses: Predictive validity of the ERASOR, PCL: YV, YLS/CMI, and Static-99. *Criminal Justice and Behavior, 36*(10), 981-1000.

<https://doi.org/10.1177/0093854809340991>

Viljoen, J. L., Mordell, S., & Beneteau, J. L. (2012). Prediction of adolescent sexual reoffending: A meta-analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99. *Law and Human Behavior, 36*(5), 423-438.

<https://doi.org/10.1037/h0093938>

Viljoen, S., Nicholls, T. L., Roesch, R., Gagnon, N., Douglas, K., & Brink, J. (2016). Exploring gender differences in the utility of strength-based risk assessment measures. *International journal of forensic mental health, 15*(2), 149-163.

<https://doi.org/10.1080/14999013.2016.1170739>

Waigandt, A., Wallace, D. L., Phelps, L., & Miller, D. A. (1990). The impact of sexual assault

- on physical health status. *Journal of Traumatic Stress*, 3(1), 93-102.  
<https://doi.org/10.1002/jts.2490030107>
- Wanamaker, K. A., Jones, N. J., & Brown, S. L. (2018). Strengths-based assessments for use with forensic populations: A critical review. *International Journal of Forensic Mental Health*, 17(2), 202-221. <https://doi.org/10.1080/14999013.2018.1451414>
- Ward, T. (2002). Good lives and the rehabilitation of offenders: Promises and problems. *Aggression and Violent Behavior*, 7(5), 513-528.  
[https://doi.org/10.1016/S1359-1789\(01\)00076-3](https://doi.org/10.1016/S1359-1789(01)00076-3)
- Ward, T., & Fortune, C. A. (2013). The good lives model: Aligning risk reduction with promoting offenders' personal goals. *European Journal of Probation*, 5(2), 29-46.  
<https://doi.org/10.1177/206622031300500203>
- Ward, T., & Mann, R. (2004). Good lives and the rehabilitation of offenders: A positive approach to sex offender treatment. In P. A. Linley & S. Joseph (Eds.), *Positive psychology in practice*, 598-616. <https://doi.org/10.1002/9780470939338>
- Ward, T., Mann, R. E., & Gannon, T. A. (2007). The good lives model of offender rehabilitation: Clinical implications. *Aggression and violent behavior*, 12(1), 87-107.  
<https://doi.org/10.1016/j.avb.2006.03.004>
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). HCR-20. *Assessing the risk of violence. Version, 2*.
- Worling, J. R., Bookalam, D., & Litteljohn, A. (2012). Prospective validity of the estimate of risk of adolescent sexual offense recidivism (ERASOR). *Sexual Abuse*, 24(3), 203-223.  
<https://doi.org/10.1177/1079063211407080>
- Worling, J. R., & Curwen, T. (2001). Estimate of risk of adolescent sexual offense recidivism

- (ERASOR; Version 2.0). *Juveniles and children who sexually abuse: Frameworks for assessment*, 2, 372-397.
- Worling, J. R., & Långström, N. (2003). Assessment of criminal recidivism risk with adolescents who have offended sexually: A review. *Trauma, Violence, & Abuse*, 4(4), 341-362.  
<https://doi.org/10.1177/1524838003256562>
- Worling, J. R., & Langstrom, N. (2006). Risk of sexual recidivism in adolescents who offend sexually. In H. E. Barbaree & W. L. Marshall (Eds.), *The juvenile sex offender* (pp. 219-247). The Guilford Press.
- Wright, R. E. (2000). Survival Analysis. In L. G. Grimm & P. R. Yarnold (Eds.). *Reading and understanding MORE multivariate statistics*. American Psychological Association.
- Yoon, D., Turner, D., Klein, V., Rettenberger, M., Eher, R., & Briken, P. (2018). Factors predicting desistance from reoffending: A validation study of the SAPROF in sexual offenders. *International journal of offender therapy and comparative criminology*, 62(3), 697-716. <https://doi.org/10.1177/0306624X16664379>
- Zeng, G., Chu, C. M., & Lee, Y. (2015). Assessing protective factors of youth who sexually offended in Singapore: Preliminary evidence on the utility of the DASH-13 and the SAPROF. *Sexual Abuse*, 27(1), 91-108. <https://doi.org/10.1177/1079063214561684>

## Vita Auctoris

NAME: Meredith Awrey

PLACE OF BIRTH: Hamilton, Ontario

YEAR OF BIRTH: 1996

EDUCATION: Saltfleet District High School, Hamilton, Ontario,  
2010-2014

University of Waterloo, Waterloo, 2014-2019 B.A.  
Hons.