EXAMINATION OF THE USE OF SELF-REPORT PSYCHOMETRICS WITHIN SEXUAL OFFENDER TREATMENT AND IN PREDICTION OF REOFFENDING

By

HELEN C WAKELING

A thesis submitted to the College of Life and Environmental Sciences of the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environmental Sciences

The University of Birmingham

September 2014

Abstract

This thesis aims to examine the utility of self-report psychometrics within delivery of sexual offender treatment. The focus is particularly on the ability of self-report psychometrics to discriminate between recidivists and non-recidivists and to predict recidivism outcome. Its findings are especially relevant to the National Offender Management Service (NOMS) who deliver sexual offender treatment across custodial and community settings in England and Wales.

Chapter 1 provides an overview of the literature on self-report psychometrics and their use within sexual offender treatment and risk assessment. Chapter 2 provides exploratory analyses into the relationship between a large battery of pre and post self-report psychometrics and recidivism outcome on a large sample of sexual offenders. Chapter 3 examines the predictive power of a selection of psychometric variables and static variables using prognostic modelling techniques. Chapter 4 examines treatment change as measured psychometrically using clinically significant change methodology and its relationship to recidivism outcome. Chapter 5 provides a summary of the previous chapters' findings and recommends further analyses and investigation. Chapter 6 attempts to generate a new shortened psychometric battery with good validity. Chapter 7 concludes the thesis with an overview, synthesis and discussion of the findings, limitations, practical implications and future research directions. The thesis found psychometrics to have limited discriminant and predictive validity, and in general static factors were better predictors of recidivism than psychometrics.

Acknowledgements

I am very grateful to my supervisors, Anthony Beech and Nick Freemantle, who have not only provided me with invaluable support, motivation and guidance over the past six years, but have done so with good humour making the process very enjoyable. They both encouraged me at all stages, inputting vitally to advice about the thesis directions, and statistical methodology. I have thoroughly enjoyed working with them on this thesis, and found the direction they provided inspirational.

I would also like to thank other people who made vital contributions to this thesis as both work and research colleagues, and friends. I have very much enjoyed working closely with Georgia Barnett on one chapter of this thesis, and appreciate the support of Ian Elliott on another chapter. I would also like to thank my previous and current Managers at National Offender Management Services, for their support and encouragement, in particular Ruth Mann, Adam Carter and Steve Webster (who inspired me to take up a career in research). Thanks also to Philip Howard, who gave me sound advice regarding pulling together a thesis, and to David Thornton who helpfully reviewed early drafts of two of the empirical chapters.

Finally my family must be thanked for their patience and encouragement over the past six years. My PhD journey has seen four house moves, and the birth of my two amazing daughters, so particular thanks go to Matt, and my girls, for their continued love and support. I'd also like to thank my father, Tony Wakeling, for inspiring me to undertake a PhD, in his footsteps, and to plough on through tough times.

Contributions

Following on from the acknowledgements, this section intends to clarify the contributions made from others on the chapters in this thesis.

Chapters one and two were conducted by myself, with support and advice from my two University Supervisors, Professor Anthony Beech, and Professor Nick Freemantle. Chapter one provides a full literature review, and presents the aims of the thesis, which I conducted myself with discussion and advice from my supervisors. Chapter two presents the first empirical study of the thesis; a preliminary examination of the predictive power of various psychometric tests used within NOMS Sex Offender Treatment Programme. I gathered the data for this study, conducted the analyses and generated the results.

Chapter three presents an empirical study of the utility of various psychometric tests and static factors in predicting recidivism using prognostic modelling techniques. The contributors to this chapter included Professor Nick Freemantle, Professor Anthony Beech, and Dr Ian Elliott, and the paper was published in Psychological Services. For this study, Professors Beech and Freemantle were involved in helping me to develop the aims and purpose of the study, and Professor Freemantle and I worked together to form the statistical plan. I gathered and organised the data, and then, with technical support from Professor Freemantle, conducted the majority of the analysis. Professor Freemantle took the lead on one aspect of the analyses, that of conducting the bootstrapping analysis to validate the final model using a statistical package R, which I did not have access to. I produced the first draft of the research paper. Professor

Anthony Beech and Dr Ian Elliott were then involved in helping to redraft the paper on receiving feedback from the journal.

Chapter four presents an empirical study examining the relationship between psychometric change, as measured by clinically significant change methodology, and recidivism outcome. The contributors to this chapter included Professor Anthony Beech and Professor Nick Freemantle, and the research was published in Psychology, Crime and Law. For this study, Professor Beech and I worked together to form the statistical plan. I gathered the data, conducted the analyses and generated the first draft of the research paper. Professors Beech and Freemantle then provided feedback in order to produce a final draft for submission. I then made changes to the paper on receiving feedback from the journal, with input from Professor Beech.

Chapter five presents a summary of the findings generated thus far from the three previous empirical chapters. The principal contributors to this chapter were myself and Georgia Barnett, a colleague from the National Offender Management Service. The paper was published in Aggression and Violent Behaviour. The purpose of the paper was to provide a policy oriented summary of previous research findings and to determine further research directions (including suggestions which led directly to the empirical study presented in chapter six). Whilst the paper presents a summary of the research presented in previous chapters of this thesis (i.e., concerning Prison data), it also provides a summary of similar work conducted by Georgia Barnett and myself (external to this thesis) on data gathered from the Sex Offender Treatment Programmes delivered in a Community Setting. This additional information was bought in to help strengthen the paper and to provide a rounded summary of the

research conducted to date in this area, in order to generate suggestions for further work required. I wrote the main bulk of the research article, with help from Georgia Barnett, specifically on those studies she led on (concerning Community data).

Chapter six presents a further empirical study which uses factor analyses in an attempt to produce a more refined and concise psychometric battery to use with sexual offenders undergoing treatment. The discriminant and predictive validity of this new battery are then examined. I gathered the data, and conducted the majority of the analyses for this study, with support and advice from both of my supervisors, Professor Anthony Beech, and Professor Nick Freemantle. Whilst I conducted the majority of the analyses for this study, Professor Nick Freemantle did contribute to helping me refine the analyses plan, and in producing the code for and conducting some of the regression analyses in this study using the computer package SAS.

Chapter seven provides the conclusions of the thesis. It attempts to draw together and discuss the findings, whilst also presenting some avenues for further research and implications for policy. I produced this final chapter by myself with support and advice from my two supervisors.

Table of Contents

Word Count by Chapter:

Chapter One	8829
Chapter Two	6123
Chapter Three	5317
Chapter Four	7281
Chapter Five	6735
Chapter Six	8632
Chapter Seven	7144
Total	50,061

List of Tables

Appendices

CHAPTER ONE

INTRODUCTION

Research Context

The need for this research arose as a direct experience of my employment with the National Offender Management Service (NOMS). I am a qualified Research Psychologist specialising in assessment and research for the Sex Offender Treatment Programmes across custody and community settings. As part of my role with NOMS I am responsible for the National self-report psychometric assessment battery used within sex offender treatment. This includes ensuring the best available assessment tools are utilised, that the methods for analysing and feeding back the assessment data are appropriate and valid, and that psychometrics and other tools are used appropriately within the risk assessment of sexual offenders. This thesis aims to examine the self-report psychometric assessments currently used within sex offender treatment, to determine how useful they are in predicting recidivism outcome, and ultimately how useful they are within the risk assessment process for sexual offenders and the evaluation of sex offender treatment. In addition, the thesis explores the possibility of improving the current psychometric assessments and makes recommendations for the future use of self-report psychometrics within sex offender treatment.

This research began in 2008, and has consequently seen a number of changes being made to the assessments used within sex offender treatment and utilises data from individuals who used previous versions of the assessment battery. The thesis attempts

to examine all of the self-report psychometrics but focuses on those which have been retained in later revisions of the assessment battery.

As a Research Psychologist working within the National Offender Management Service I have been fortunate to be able to work with large National datasets, and many of the research studies within this thesis use these large datasets of those who have undertaken the Sex Offender Treatment Programme in custody. Additionally, I have been able to access networks of other professionals, with whom I have formally and informally shared this research. This has been an important process for me, and has enabled me to reflect on research questions, methodology and to formulate and develop each successive chapter taking into account advances and other emergent research in the field. Having worked in this area for a number of years, I have also been aware of my potential bias in the use and validity of self-report psychometrics. I have attempted to address this by seeking feedback via the supervision process, and by publishing the empirical chapters where possible to ensure objective views are obtained from professionals and researchers working in this field but who are not directly employed by NOMS.

Thesis Structure

This thesis uses quantitative research methodology and draws data from the National population of sexual offenders who have undertaken the Sex Offender Treatment Programme in custody between 1996 and 2010. An overview of the chapters follows:

Chapter One – Introduction. This chapter presents an introduction to the nature of the thesis and provides an outline of the structure of the thesis. The main body of the

chapter presents a literature review of the relevant literature to the thesis. This includes background to the treatment of sexual offenders, and information on self-report psychometrics, and how they are used within sex offender treatment and the risk assessment of sexual offenders. The literature review concludes with a rationale for the research in the thesis.

Chapter Two – The relationship between self-report psychometrics and recidivism amongst sexual offenders: An examination of individual psychometrics and psychometrics grouped into risk domains. This chapter provides the initial exploratory analyses of the relationship between the pre and post-treatment self-report psychometrics and sexual and violent recidivism. A variety of analytical techniques are used to examine this relationship, and all of the self-report psychometrics used in the period of data collection were examined. The analyses reveal the discriminant and predictive validity of the individual psychometrics.

Chapter Three – Identifying predictors of recidivism in a large sample of UK sexual offenders: A prognostic model. This chapter expands on the analyses presented in chapter two (using the same sample), and provides a more refined analysis of the predictive power of a selection of the pre and post-treatment self-report psychometrics used within sex offender treatment. Prognostic modelling techniques are used to examine how well psychometric variables can predict sexual and violent recidivism alongside static variables. This research study was accepted for publication in a peer reviewed journal, and as such the article is presented as the main body of this chapter.

Chapter Four – An investigation into treatment change and its relationship to recidivism in a sample of 3773 sex offenders in the UK. Previous chapters have examined the predictive power of self-report psychometrics at the pre and post-treatment stages. This chapter furthers this work by examining the predictive power of *change* as measured by self-report psychometrics, using clinically significant change methodology, using the same sample as chapters two and three. This research study was accepted for publication in a peer review journal, and as such the article is presented as the main body of this chapter.

Chapter Five – A summary of the relationship between psychometric test scores and reconviction in sexual offenders undertaking treatment. This chapter presents a summary of the research conducted so far in the previous chapters, and synthesises some emergent research in the field, in order to generate directions for the final chapters of the thesis. It was useful to examine the findings thus far in this way in order to reflect on the progression and journey of the thesis. This summary was accepted for publication in a peer review journal, and therefore the article provides the main body of this chapter.

Chapter Six – Development of a new set of psychometric measures for risk of sexual offending. This chapter provides an attempt to produce a new battery of self-report psychometrics to be used within sex offender treatment which is more concise than the previous battery, and which has greater predictive power. Factor analyses are used to attempt to produce new psychometrics in this way, and then an examination of the new psychometrics' discriminant and predictive validity is conducted.

Chapter Seven – Discussion and recommendations. The final chapter presents a synthesis of the empirical evidence presented and assimilates these results into other research findings, in order to generate conclusions and recommendations for clinical practice and further research. The use of psychometrics within the treatment and risk assessment of sexual offenders will be summarised, and suggested directions for the use of psychometrics within the sex offender field will be presented with consideration of the findings from the empirical research conducted in the previous chapters of the thesis.

Sexual Offending and Treatment Provision for Sexual Offenders

Sexual offending is a serious and widespread public health concern (Furby, Weinrott & Blackshaw, 1994; Mann, Hanson & Thornton, 2010), which has enormous social, physical and economic consequences (Finkelhor, 1984). The consequences do not merely extend to the victims, but to society as a whole. Thus, there is a need for effective and constructive treatment for sexual offenders (Friendship, Mann & Beech, 2003). Treatment interventions or Offending Behaviour Programmes (OBPs), which aim to reduce rates of future reoffending, are now routinely part of correctional services for sexual offenders, and go some way to address the societal problem of sexual offending.

The most recent evidence suggests that programmes for offenders are most likely to be successful in achieving reductions in recidivism if they are founded on the Risk-Need-Responsivity (RNR) model of offender rehabilitation (Andrews & Bonta, 2010). That is, treatment should be matched to risk of reoffending, it should be focused on criminogenic needs, and should be delivered in a responsive way. Programmes which

follow the RNR principles have been shown to lead to the largest reductions in recidivism (Hanson, Bourgon, Helmus & Hodgson, 2009). In recent years, the Good Lives Model (GLM; Ward, 2002; Ward & Brown, 2004; Ward & Marshall, 2004) has also been incorporated into treatment models for sexual offenders, which assimilates a focus on an individual's values and strengths as well as risk factors for sexual offending.

NOMS Sex Offender Treatment Programmes

The NOMS Sex Offender Treatment Programmes (SOTPs) are a suite of cognitive-behavioural programmes, which adhere to the RNR principles and also incorporate aspects of the GLM. For a comprehensive review of NOMS SOTPs please refer to Mann and Thornton (1998), Mann and Fernandez (2006), and Barnett and Wakeling (in press). The SOTP Core programme, the longest standing of the SOTPs and the main treatment sample for this thesis, has been delivered in prisons since 1991 and is currently provided in 25 prison establishments across England and Wales. The programme was designed for offenders who have committed at least one sexual offence or an offence with a sexual element, who are medium to very high risk of further sexual reconviction, who are not in categorical denial, and who are willing to engage in the treatment process. The main focus of the programme is to help the offender develop meaningful life goals that will lead him away from offending, giving him the opportunity to practice the skills necessary to achieve these new goals.

The Rolling programme is targeted at low risk male sexual offenders. The programme covers the same topics as the Core SOTP but provides a milder level of treatment with more emphasis on relationships skills and attachment style deficits. The Core and

Rolling programmes are the first stage programmes for all sexual offenders with an IQ of greater than 80. The other SOTPs are second stage programmes which include the Better Lives Booster programme, the Extended programme, and the Healthy Sexual Programme, which offenders would do if further work is required having completed either the Core or the Rolling programme.

The only outcome study which has been conducted on the NOMS SOTPs to date was conducted by Friendship et al. in 2003. The study found the Core programme to be effective at reducing sexual and violent recidivism for medium risk sexual offenders. A sample of 647 treated offenders were matched with a sample of 1910 male offenders who had not taken part in the programme and the treated offenders had statistically significantly lower sexual and/or violent reconviction rates at two years than the untreated offenders. The programme was not sufficient to reduce recidivism in high risk sexual offenders, and the low risk offenders had very low levels of reconviction whether treated or not.

Further evaluations of the SOTPs have not been possible mainly due to the difficulties in designing and executing robust methodologies (see Hollin, 2008; Marshall & Marshall, 2007). However, the evidence gathered thus far indicates that NOMS SOTPs provide treatment to sexual offenders which adhere to the principles of effective intervention.

Risk Assessment of Sexual Offenders

Offender risk assessment involves examining the behaviour of offenders in order to establish statistically significant differences between those who offend at high rates

and those who offend at low rates. Reoffending prediction research therefore creates methods to identify whether an individual is at high or low risk of going on to reoffend. Information about an individual's risk can be used to assist treatment planning, and can inform incarceration and/or release decisions.

Types of Risk Factors

There are two main types of risk factors; static and dynamic (e.g. Andrews & Bonta, 2006). Static risk factors are relatively fixed, cannot be deliberately changed and have a reliable relationship with offending. Examples include age and criminal history. Static risk factors are often pooled together to produce actuarial risk assessment tools, which are statistically derived tools designed to predict the likelihood of a future behaviour. In terms of sexual offender risk assessment these tools put individuals into risk groups who differ reliably in their rates of sexual reconviction. Actuarial risk assessment tools in general are good predictors of reconviction (Hanson & Morton-Bourgon, 2009). However, due to the fact that they are based on fixed, static risk factors, actuarial tools cannot tell us whether a psychological or behavioural change on other *changeable* risk factors is meaningful in terms of reducing or increasing the chances of further offending.

Dynamic risk factors are changeable and are therefore potentially amenable to treatment. Dynamic risk factors are also sometimes described as *criminogenic needs*. Hanson (1998) described dynamic risk factors as those which are related to recidivism, have the potential to change, and are associated with a reduction or increase in recidivism rate when changed. Dynamic risk factors are often pooled together into structured professional judgement (SPJ) or guided clinical judgement

tools to provide frameworks with which to assess sexual offenders. Within such tools, the judgement of clinicians is aided by guidelines designed to support clinical judgement (Tully, Chou & Browne, 2013). Whilst the overall decision on risk level is up to the assessor, the tools include factors which have been empirically shown to be related to sexual recidivism. Examples include the Sexual Violence Risk-20 (SVR-20: Boer, Hart, Kropp & Webster, 1997), and the Risk for Sexual Violence Protocol (RSVP: Hart, Kropp & Laws, 2003). The term mechanical tools (Hanson & Morton-Bourgon, 2009) is used to describe those tools which have a defined set of items, and a defined method for producing an overall risk level.

Dynamic Risk Factors for sexual offenders

Mann, Hanson and Thornton (2010) conceptualised dynamic risk factors as psychologically-meaningful propensities, which may or may not manifest themselves within a given time period, that lead to predictable expressions of thoughts, feelings and behaviour. These propensities must have a plausible reason for being considered to be causal of sexual offending, are amenable to change, and are therefore able to be targets of treatment designed to reduce reoffending. For example, Mann et al. (2010) suggest that it is an individual's propensity to be drawn towards a criminogenic environment that is the reliable, long-term, indicator of risk of reoffending, rather than the criminogenic environment itself. Mann et al. (2010) reviewed the evidence for the predictive validity (factors' ability to predict sexual recidivism) of psychologically meaningful risk factors, and in doing so identified those risk factors which are empirically supported, those which are promising, those with are unsupported or worth exploring, and those which are not risk factors for sexual offending. This work

provides a good basis for identifying those risk factors of sexual offending which are empirically supported.

The Structured Assessment of Risk and Need (SARN)

SOTP utilises a framework of static and dynamic risk factors, or a 'mechanical tool' which can be referred to as the 'Structured Assessment of Risk and Need' (SARN, formerly SRA, Thornton, 2002; Webster et al., 2006). The SARN is a comprehensive three-stage process for assessing risk, need and progress in sexual offenders. SARN uses a combination of actuarial assessment (RM2000/s) and structured professional judgement (examination of dynamic risk factors), to provide a comprehensive assessment of risk of sexual reoffending. A combined approach is recognised as the most valid (in prison samples, Beggs & Grace, 2011; Knight & Thornton, 2007; Olver, Wong, Nicholaichuk & Gordon, 2007, and in community samples, McGrath, Lasher & Cumming, 2012), practical and ethical approach to risk assessment (Barnett & Mann, 2011). SARN is intended primarily to focus treatment planning and the evaluation of progress on those factors that have an established relationship with sexual reoffending (as defined by Mann et al., 2010).

The first stage involves examination of static risk using the actuarial risk assessment tool, Risk Matrix 2000/s (Thornton et al., 2003). This initial stage guides level of treatment provision in accordance with the RNR principles and determines which of the family of SOTPs individuals are recommended to embark on. The RM2000/s has been shown to have good predictive power (e.g. Barnett, Wakeling & Howard, 2010).

The second stage involves an assessment of dynamic risk factors, as is often referred to as the Treatment Needs Analysis (or TNA). Within SARN, the term *treatment needs* is used to describe dynamic risk factors, and within this thesis the two terms are used interchangeably. This second stage of the SARN uses a structured professional judgement framework. That is, it provides an assessment of a series of risk factors which are specified in advance but the overall evaluation of risk relies on the evaluator. Within the TNA, the clinician provides a rating for whether each risk factor was strongly characteristic (2), mildly characteristic (1) or absent (0) in the lead up to the offence (offence chain) and in the offender's life generally (generality). Clinicians use a variety of information to make this judgement. The dynamic risk factors within SARN are used to guide treatment, and to measure individual treatment change and progression. All of the SARN dynamic risk factors have been shown to have a reliable relationship (or promising relationship) with risk of recidivism (Mann et al., 2010). Treatment intends to target these dynamic variables in an attempt to reduce factors contributing to sexual offending.

The SARN identifies 15 main dynamic risk factors for sexual offending, which are adult characteristics, and can be thought of as psychological propensities that predispose someone to committing a sexual offence. These 15 risk factors cluster into four main dynamic risk domains: sexual interests, offence supportive attitudes, poor interpersonal functioning or socio-affective functioning, and poor self management (Hanson & Harris, 2001; Mann & Fernandez, 2006; Mann et al., 2010; Thornton, 2002). These domains and corresponding risk factors are described in detail below.

Domain 1: The sexual interest domain

This consists of risk factors relevant to offence related sexual interests including:

- Sexual preoccupation, which relates to the intense interest in sex that tends to dominate psychological functioning.
- Sexual preference for children, referring to an intense interest in sexual activity with children.
- Preferring sex to include violence which comprises two aspects: Being preferentially sexually aroused to the idea of forcing sex (coercive rather than consensual sexual activity); or being sexually aroused to the idea of inflicting violence, pain, terror, humiliation, destruction or exercising abusive control over another person.
- Any other offence related sexual interest refers to any sexual interest not covered by the above categories which is related to the offending. It is intended to pick up other, unusual sexual interests, which appear to be more easily gratified through sexual offending than through legal sex.

Domain 2: The offence supportive attitudes domain

This contains four risk factors related to distorted thinking in sexual offenders, including:

- Adversarial sexual attitudes or believing that men should dominate women. These attitudes involve three closely linked components: 1) Believing that men need to be tough and dominating while women should be submissive and accepting; 2) Seeing sexual encounters between men and women as essentially adversarial; and 3) Seeking sexual gratification from women without any empathic concern for the woman's experience.

- Believing one has the right to sex, which involves an excessive sense of entitlement taking two primary forms: a sense of sexual entitlement and a generalized sense of entitlement.
- Child abuse supportive beliefs include three main types of belief: 1) Children are little adults with an adult capacity to enjoy sex and to consent to it; 2) Children can be sexually knowing and sexually provocative, deliberately inviting sexual interest from adults; and 3) Sex with children is harmless for the child, or is positively beneficial, so long as the man is gentle. These beliefs make it easier for an offender to enjoy sexual fantasy about children and permit him to overlook signals that a victim is distressed during an offence.
- Believing that women can't be trusted. The core of this belief is the idea that women are generally deceptive and malicious towards men.

Domain 3: The socio-affective functioning domain

This consists of four risk factors relating to deficits in the skills required to develop and maintain successful adult relationships (Cortoni & Marshall, 2001). Specific risk factors within this domain include:

- Inadequacy, which involves a combination of low self esteem, subjective feelings of loneliness, and having an external locus of control.
- Feeling more comfortable with children than adults. This relates to an offender who has a greater emotional congruence with children. This factor suggests a problem with adult relationships combined with a pleasure in relationships with children, and involves emotional needs (not just sexual needs).
- Being suspicious and angry with others. This relates to grievance thinking, which is defined by having difficulty seeing others' point of view, believing that others

have wronged you and are likely to do so again, having angry rumination over past wrongs, having a sense grievance against the world and others, and vengefulness.

- Not having intimate relationships with adults. This refers to a relative absence of emotionally intimate marital-type relationships (either homosexual or heterosexual). It does not refer to the closeness of family relationships or social friendships even though some of the skills needed for an emotionally intimate marital relationship can be practised or displayed in these other relationships.

Domain 4: The self management domain

This consists of three risk factors which relate to poor self regulation, including:

- Impulsivity, which relates to a lifestyle dominated by impulsive irresponsible decisions, driven by the need for stimulation, and not organised by realistic long-term goals.
- Poor cognitive problem solving. This includes both failing to deploy cognitive skills or using faulty (maladaptive) coping strategies to deal with problems. This is about offenders characteristically getting into difficult situations in which they cannot manage the situation or their emotions, rather than deliberately intending to do or cause bad things.
- Not being able to control one's emotions. Chronically poor emotional control involves repeatedly behaving in an emotional, unconstrained manner. Poor emotional control often displays as uncontrolled outbursts of emotion.

The third stage of SARN involves assessing an individual's progress in treatment against the targets set in the form of a risk report. It provides an assessment of overall

level of risk of serious harm, based on sections one and two of the SARN, an assessment of current functioning and progress in treatment, any idiosyncratic factors relevant to an individual (e.g., psychopathy, responsivity issues, mental health issues), and factors that could protect someone from reoffending, or make desistance from crime more likely.

Research Evidence for the SARN

The SARN TNA has been found to be predictive of sexual reconviction with offenders under community supervision (Craig, Thornton, Beech, & Browne, 2007), with prisoners participating in treatment (Thornton, 2002), and with sexual offenders being assessed for a civil commitment program (Knight & Thornton, 2007). In the original development and validation studies Thornton (2002) found the SRA to have an AUC of 0.78. The limitations to these studies include the fact that only psychometric data were used to assess the presence of the dynamic risk factors and that only three of the four SRA domains were examined (there were unfortunately no psychometric measures utilised to assess the Sexual Interests domain). Recently Thornton and Knight (2013) scored the SRA on a sample of 566 adult men who, between 1959 and 1984, were considered for a civil commitment in the United States. In this study the attitudes domain was not scored on the basis that accurate unbiased self-report data was not available. The other three domains were scored by SRA raters. At the five year follow-up the SRA demonstrated good predictive validity for sexual recidivism (AUC = .73), and at 10 years the AUC was .72. Analyses revealed that the overall SRA score added significant incremental validity to the RM2000/s.

Research shows that empirically-derived actuarial assessments are significantly better than unstructured clinical judgement, at predicting sexual, sexual or violent and

general offending (Hanson & Morton-Bourgon, 2009). Structured professional judgement (which stage 2 of the SARN process fits into) also fared better than unstructured clinical judgement, but were slightly less predictive than actuarial tools. This study also found that risk assessments which included dynamic as well as static factors were more accurate at predicting sexual recidivism outcome, than those that were based on static factors alone.

Although there is thus a body of supporting evidence for the SARN, the studies above are limited by relatively small sample sizes and other methodological issues. For this reason, the TNA is not currently used to predict risk of sexual reoffending. The primary purpose of the TNA is to identify those psychological propensities related to risk of sexual reoffending, that are relevant to the individual being assessed.

Methods for determining presence of risk factors

A variety of different methodologies have been employed to measure dynamic risk factors (Craissati & Beech, 2003) in the sex offender literature. These include examining information obtained from self-report psychometric measures at pre and post-treatment (e.g., Hudson, Wales, Bakker & Ward, 2002), using subjective ratings of presence made by clinicians (e.g., Olver, Wong, Nicholaichuk & Gordon, 2007), or using a triangulation of information obtained from psychometrics, observation and interview. It is important to note the often different methods used within research studies and those methods used for the purpose of routine assessment of sexual offenders in clinical settings. For the purpose of the routine assessment of sexual offenders in a practical setting, NOMS utilise the latter, a triangulation of information

to determine presence or absence of the 15 SARN dynamic risk factors or treatment needs within sexual offenders.

Methods for determining within-treatment outcome

Determining treatment change is a related, but different concept. Similarly to the identification of treatment need, treatment change or within-treatment outcome is determined using different methods across research and clinical settings, and within treatment change can be examined using a variety of different methods (Beggs, 2010). Whilst some studies use pre-post psychometric change (e.g., Hudson et al., 2002), others use tools developed specifically to assess treatment progress such as the Goal Attainment Scale (Hogue, 1994), or the Treatment Readiness Scale (Serin & Kennedy, 1997), some make professional clinical judgments of treatment progress (e.g., Looman, Abracen, Serin & Marquis, 2005), and others still evaluate changes in risk through the use of structured dynamic risk assessment tools (e.g., Olver et al., 2007; Wong & Gordon, 2006). For the purpose of the routine assessment of sexual offenders in clinical settings, NOMS utilise the SARN, as described above, which uses a triangulation of evidence, including psychometric information, and clinical assessment to determine the progress made by individuals and changes that have been made as a result of treatment.

Self-report Psychometric Measures

Psychometrics is a field of study relating to the development and refinement of theoretical approaches to measurement and the construction of instruments and procedures for measurement (Michael & Li, 2010). Self-report psychometric

measures are used to tap into unobservable latent variables, and measure knowledge, abilities, attitudes and personality traits. Classical test theory (CTT) is the most commonly applied psychometric theory in the field of psychology. It relates to the breaking down of an individual's observed score on a test into a true score component, that is free from internal or external error. The aim of CTT is to understand and improve the reliability of psychological tests, by examining one source of error at a time. A more recent psychometric theory or set of theories can be referred to as Item Response Theory (IRT), which examines the pattern of item responses on psychometric measures (Nunnally & Bernstein, 1994). IRT applies mathematical models to data from psychometric measures in order to measure latent variables. IRT assumes that there is a correlation between an individual's response on one item and their overall ability on the latent trait which is being measured. The most basic IRT model therefore assumes that items vary only with respect to their difficulty. Whichever theory is utilised, researchers and developers of psychometric tools must demonstrate that tools are reliable and valid.

Reliability and validity are considered by many to be the most important aspects of a psychometric test (Rust & Gollombok, 1999). Reliability refers to the accuracy of the tools or the dependability of a measure over time. Reliability can be measured in four main ways: 1) The internal consistency of a scale examines the extent to which the items are assessing the same characteristic or latent variable. It measures the precision between items of a scale by assessing the inter-correlation between items. If the items were measuring the same overall construct, then scores on each of the items would be expected to correlate both with each other and the overall scale. The general preferred measure of internal consistency is Cronbach's alpha (Cronbach,

1951; Kline, 1993). 2) Test-retest is another form of reliability, which scrutinizes the stability of the tool over time. If the correlation between scores at two different time points is high, then the consistency of the tool is regarded as high. 3) Split half reliability estimates reliability based on the correlation of two equivalent forms of a measure. 4) Inter-rater reliability examines the correlation of scores between or amongst two or more raters who rate the same item, scale or tool. There are many guidelines for interpreting reliability estimates. For research purposes, Nunnally and Bernstein (1994) suggest that 0.70 is a sufficient reliability estimate. However others have suggested that if scores are being used for clinical purposes, a reliability of 0.85 or higher should be required (Rosenthal & Rosnow, 1991).

Validity refers to the ability of the tools to measure what they intend to measure. Reliability is a necessary but not sufficient condition for validity; that is measures cannot be valid without being reliable. The main forms of validity are as follows: 1) Face validity refers to an assessment of whether the tool appears to be a credible and plausible measure of the construct. 2) Content validity refers to the ability of the tool to adequately sample the content in the domain of interest in relation to theory. Face and content validity both involve examination of the tool's items and assessing their suitability qualitatively in terms of what they appear to measure (face), and according to the theoretical background of the construct (content). 3) Construct validity is demonstrated by convergent and discriminant validity. Convergent validity is the extent to which the construct of interest and other constructs, which are deemed to be theoretically related, are observed to be so. Discriminant validity on the other hand examines the extent to which the construct of interest is related to constructs, which are assumed to be largely independent, or its ability to discriminate between groups

who should score differently on the construct. 4) Criterion validity evaluates the measure by comparing it with some other standard. There are two types of criterion validity, concurrent and predictive. Concurrent validity is evaluated by comparing the psychometric test with another test that already exists (i.e., another measure of the same latent variable), and predictive validity is the extent to which the construct can predict some type of criterion to be tested in the future (e.g., recidivism).

Use of Psychometrics within SOTP

Psychometric measures are used in forensic settings to measure latent variables. Within offending behaviour treatment programmes, they are used primarily to measure those dynamic risk factors which are targeted on those programmes. They are usually given both prior to the treatment starting, and at the end of the programme. Often the psychometrics are used for evaluation purposes, to determine the short-term impact of treatment programmes delivered to offenders in prison and community settings (see Wakeling & Travers, 2010 for a review). They are also sometimes used to help identify areas of treatment need and to help determine individual treatment change.

Self-report psychometrics are currently administered pre and post-treatment on the NOMS SOTPs. The self-report psychometrics are used to help assess the presence or absence of the 15 SARN risk factors described previously, along with other evidence such as the TNA interview, treatment participation, review of treatment programme products, offence descriptions, behaviour monitoring and observation, and file review. Within NOMS SOTP, an individual's pre and post-treatment scores for each

psychometric are plotted on a psychometric profile graph and compared with a normative group of low risk sexual offenders. All scores are converted to Z scores and then standardised so that you can see an individual's score on each psychometric relative to the normative group. Any score that is more than one half of a standard deviation away from the mean is regarded as a clinically meaningful difference from the average score. This pre-post psychometric change (observed on these profiles) contributes to an assessment of change or progress in treatment within the SARN framework.

SARN authors all attend training (and must meet minimum standards to pass this training) in order to complete SARN assessments, and this training stipulates that psychometric data should not be relied upon in isolation to assess presence or absence of a dynamic risk factor or to assess progress in addressing that factor. These decisions must be made on the basis of a range of evidence. However, much of the research base utilises studies which identify the presence or absence of risk factors using one method alone. The research within this thesis relates to the use of self-report psychometrics *alone* in measuring the dynamic risk factors, and whether they are reliable contributors to the process currently used by NOMS. In order to have confidence in the use of such assessment measures (within research and/or clinical practice), researchers and clinicians need to be sure that the measures are both valid and reliable, are measuring what they purport to measure, and are related to the outcome of interest (Grady, Brodersen & Abramson, 2011).

Since psychometrics are used to help identify treatment need and treatment change within SOTP, and to contribute to an offender's risk assessment, then they must have

good discriminant and predictive validity. That is, they must have a demonstrated relationship with the main outcome of interest within sex offender treatment, future reconviction. Decisions about treatment need and progress may have a direct impact on an individual's progression through the criminal justice system, and therefore there are ethical considerations regarding the selection of tests, their reliability and validity, and their use within sexual offender treatment and risk assessment.

In a recent review of the use of psychological measures with adult sexual offenders conducted by Grady et al. (2011), the authors suggest that within sex offender treatment provision, often there is a big gap between the main outcome of interest (recidivism) and the proximal outcome (changes in dynamic risk factors). They suggest that we need to focus our efforts on understanding treatment change, using measures which are reliable and valid, and which are effective measures of specific areas identified as dynamic risk factors (Fanniff & Becker, 2006). The use of valid psychometric tests to do this will enable researchers and clinicians to determine whether sexual offenders who complete treatment are demonstrating changes in the core treatment need areas addressed on the programme. In turn, this also presents a practical alternative for evaluating treatment programmes (Beggs, 2010), the problems of which have been well documented elsewhere (e.g., Hollin, 2008). It could be, for example, that pairing short term outcomes such as psychometric change, with long term reconviction may be a more accurate method of examining the impact of a treatment programme (Friendship, Falshaw & Beech, 2003).

Marques et al. (2005) who conducted a randomised controlled trial of a sex offender treatment programme, concluded that there is a need for 'standardised, empirically-

validated measures of treatment progress' (p. 103) in order to make inroads in understanding how short term change may be related to long term recidivism outcome. Whilst we know that there are static, historical factors and dynamic risk factors associated with recidivism risk (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2005; Lusignaan & Marleau, 2007), less is known about the characteristics that need to *change* in order to reduce risk, or whether dynamic risk as measured by psychometric measures are related to recidivism outcome. Evidence must still be gathered to confirm how best to measure dynamic risk factors, and to confirm the notion that changes on dynamic risk factors (as measured psychometrically) are associated with meaningful changes in offenders' recidivism risk (Beech, Fisher & Thornton, 2003).

The practical advantages of using self-report psychometric measures include the fact that they are objective, and economical to administer. They also provide a convenient method of gathering information with minimal possibility of assessor biases. Although some research has found social desirability not to be an issue with self-report measures within general populations (e.g., Ones, Viswesvaran & Reiss, 1996; Richman, Kiesler, Weisband & Drasgow, 1999), other researchers criticise their use and suggest they are particularly prone to faking, lying and reporting in a self-enhancing manner within offender populations (e.g., Holden, Kroner, Fekken & Popham, 1992; Schretlan & Arkowitz, 1990; Tierney & McCabe, 2001). Self-report measures also tend to rely on individuals possessing a level of personal reflection and insight, which may not always be present in an offender population (Mathie & Wakeling, 2011). Offenders might be more likely to self-deceive, to protect their personal identities, and to reduce their feelings of shame and guilt about their

offending. In fact, many of the general problems often associated with the use of psychometrics may be more pronounced in an offender population, due to the fact that an offender's progression through the criminal justice system may, in part, depend on the outcome of a set of psychometric measures. In a review of the available psychometric measures to use with sexual offenders and measure appropriate dynamic risk factors, Grady et al. (2011) conclude that there are few valid and reliable instruments available to researchers and clinicians to accurately measure sex offenders' deficits areas. They call explicitly for further research to identify specific instruments or tools that are validated with sexual offenders that can be used both to better understand the dynamic risk factors of sexual offenders, and have the greatest impact on the overall recidivism rates of treated offenders.

The psychometric measures used within NOMS SOTP are described in chapter 2, along with their reliability and validity estimates. The majority of these measures have been validated for use with sexual offenders, and have demonstrated reliability and validity. However, most of the psychometric measures do not have an estimate of their discriminant or predictive validity; that is their relationship with recidivism outcome. The present thesis attempts to redress this omission.

Research Findings on the Predictive Validity of Self-report Psychometrics

Within the literature generally there is support for the notion that psychometrics can detect a positive change (or reduction) in dynamic risk following treatment (e.g., Beech & Hamilton-Giachritsis, 2005; Bickley & Beech, 2003; Bakker, Hudson, Wales & Riley, 1998; Hudson et al., 2002; Marshall et al., 2002; Marshall et al., 2003; Nunes, Babchishin & Cortoni, 2011; Quinsey, Khann & Malcolm, 1998; Wakeling,

2007; Webster, Mann, Thornton & Wakeling, 2007), when the psychometrics are empirically related to risk, and are targeted in the treatment programme (Beggs, 2010), and using a variety of methodology to assess change. This is promising, but in order to be confident that these reductions are meaningful, their relationship with recidivism (or their predictive validity) must be established. The literature is less consistent when investigating this link between psychometric scores and recidivism.

Studies examining the relationship between pre and/or post-treatment scores and recidivism

Some studies have found a poor relationship between psychometrics and treatment outcome (e.g., Hanson, Cox & Woszczyn, 1991; Hanson, Steffy & Gauthier, 1993; Hanson & Wallace-Capretta, 2000). In the non sex offender specific literature, Bowen, Gilchrist and Beech (2005) found no differences in pre-treatment measures of anger, locus of control, self-reported abusive behaviours, and pro-offending attitudes between domestic violent reoffenders and nonreoffenders. Wilkinson (2005) further found no associations between recidivists and non-recidivists on a battery of self-report measures on a small sample of Reasoning & Rehabilitation programme completers. However, the use of outdated or non-standard tests, or tests not specifically designed/validated with offenders, may limit the generalizability of these findings (Proulx et al., 1997). Additionally, self-report measures would need to reflect the constructs thought to be related to risk factors. Walters (2006) conducted a meta-analysis examining the ability of various self-report measures in predicting general recidivism, institutional misconduct, and violence. The findings suggest that self-report psychometrics can predict outcome as well as static risk assessment tools, but only if these tools are measuring constructs associated with criminogenic needs

and empirically related to risk. Amongst sexual offenders, some studies have also found psychometrics to have poor predictive validity. Proulx et al. (1997) for example, examined a whole battery of psychometric tests, none of which were able to discriminate between recidivists and non-recidivists.

Other studies, however, have found a link between level of dynamic problems, as measured by self-report psychometric measures, and recidivism outcome. In the general offender population, the Self-Appraisal Questionnaire (SAQ), a self-report questionnaire with six subscales measuring a range of criminal attitudes and behaviours, has shown good predictive validity in a series of studies (Kroner & Loza, 2001; Loza et al., 2004; Loza & Loza-Fanous, 2000, 2003; Mills, Loza & Kroner, 2003). Andrews, Kiessling, Mickus and Robinson (1986) also found that self-report questionnaires could add incremental validity to the Level of Service Inventory (LSI) in predicting recidivism in a general offender population.

Relationships have also been observed in the sexual offender population (e.g., Beech, Friendship, Erikson & Hanson, 2002; Craig et al., 2007; Raynor, 1998; Thornton, 2002). Beech's deviancy index (Beech, 1998) has been widely used in Sex Offender programmes delivered in the community in England and Wales. This construct puts offenders into high or low deviancy groups based on how much they differ on psychometric measures from a non-offending population. Beech et al. (2002) found that adding this measure of deviance significantly increased the accuracy of risk prediction beyond the level achieved by an actuarial static measure.

In the original development and validation studies for the SARN, Thornton (2002) found that repeat sexual offenders scored in a significantly more dysfunctional way on a set of psychometrics than offenders with only one sexual conviction. In a second study, Thornton (2002) examined a sample of 117 imprisoned sexual offenders. Offenders were classified as low, moderate or high deviance according to their scores on a set of pre-treatment psychometrics. None of the 40 low deviance men, five percent (2/43) of the moderate deviance men, and 15 % (5/34) of the high deviance men were sexually reconvicted, which produced an AUC of .78. Although the static risk predictor (Static-99) was more accurate than the deviance ratings, the Static-99 and the deviance ratings independently predicted sexual reconviction. The combination of static and dynamic risk led to a significant improvement in prediction. Craig et al. (2007) found that sexual offender recidivists scored significantly worse on the Multiphasic Sex Inventory (MSI; Nichols & Molinder, 1984) and the Special Hospital Assessment of Personality and Socialisation (SHAPS; Blackburn, 1982) than non-recidivists. More recently, a meta-analysis of measures of cognitive distortions (Helmus, Hanson, Babchishin & Mann, 2013) found that some measures including the Abel-Becker Cognition Scale (Abel et al., 1989) and the Children and Sex Cognitive Distortions Scale (Beckett, 1987) showed small significant associations with sexual recidivism. Allan, Grace, Rutherford and Hudson (2007) used a four-factor dynamic risk framework incorporating a battery of psychometric tests, and found that two of the four risk domains had independent correlations with sexual recidivism, after controlling for static risk and pro-offending attitudes amongst sexual offenders.

Many of these studies have only examined pre-treatment psychometric scores. But if treatment works, and the pre and post-treatment psychometrics are valid measures of

risk factors, then post-treatment scores should perhaps be more strongly associated to outcome than the pre-treatment scores. Of those who have looked at both pre and post-treatment scores and their relationship to recidivism amongst sexual offenders however, the evidence seems to suggest that pre-treatment scores may be better predictors than post-treatment scores (e.g., Hanson & Wallace-Capretta, 2000). Quinsey et al. (1998), for example, found that pre-treatment Buss Durkee assessment scores (Buss & Durkee, 1957) were better predictors of recidivism than post-treatment scores.

Studies examining the relationship between treatment change and recidivism

There are a number of studies which have examined "within-treatment" change using judgements of change made by clinicians, or using structured rating systems without the addition of psychometric data (e.g., Barbaree, Seto, Langton & Peacock; 2001; Olver et al., 2007), or using simple ratings of change (e.g., Scalora & Garbin, 2003; Looman et al., 2005), again with mixed results. Important for this thesis are those studies which have examined psychometric treatment change scores and their relationship to recidivism. Whilst some of these studies indicate that treatment change as measured by psychometrics can be related to recidivism outcome (e.g., Beech, Erikson, Friendship & Ditchfield, 2001; Hedderman & Sugg, 1996; Hudson et al., 2002; Marques et al., 2005), others have found pre to post-psychometric change to be unrelated to recidivism (e.g., Hanson & Wallace-Capretta, 2000; Wilkinson, 2005).

Serin, Lloyd, Helmus, Derkzers and Luong (2013) provide a comprehensive review of studies linking individual change scores to recidivism in areas of cognition, violence and substance misuse. Of the 49 studies examined, changes were shown to relate to

recidivism in areas of antisocial attitudes, antisocial beliefs, antisocial personality, social support and substance misuse. The authors conclude that although there are promising links between change and recidivism, further high-quality studies are required. Beggs (2010) conducted a similar review of within-treatment outcome studies among sexual offenders, which concludes that the evidence amongst this group of offenders is somewhat mixed.

Bakker et al. (1998) and Hudson et al. (2002) both report on the Kia Marama treatment programme for sexual offenders in New Zealand. Bakker et al. (1998) compared the pre and post-treatment psychometric scores of recidivists and non-recidivists, computing an effect size to compare the groups' response to treatment. They found significant differences on some measures, such as the Trait Anger and Anger Suppression subscales of the State-Trait Anger Expression Inventory (STAXI; Spielberger & Sydeman, 1994), and the Perspective Taking subscale of the Interpersonal Reactivity Index (IRI; Davis, 1980); with reoffenders scoring worse than non-recidivists. For the majority of the psychometric measures in the assessment battery, however, the effect was not significant. Although the recidivist group was small, this is not a promising result for the predictive validity of psychometric measures. Hudson et al. (2002) calculated individual change scores on each psychometric, and observed the correlation between these and recidivism. Although the correlations were generally in the desired direction, few were significant. The measures targeting interpersonal difficulties were generally the most promising, with assertiveness and empathy scores related to recidivism. However, positive change on some measures was actually related to increased sexual recidivism. One criticism of the method Hudson et al. (2002) used was that they did not take into account the pre-

treatment level, and as such those who are most likely to show change are probably those who are most deviant and have the worst scores to begin with.

Beech and Ford (2006) used clinical significant change methodology to determine the relationship between treatment change and recidivism outcome in a sample of 51 sexual offenders. None of those deemed to have responded to treatment as measured by the Jacobsen method of clinical change (Jacobsen & Truax, 1991) were reconvicted, whereas at two years 14% of those who were not deemed to have changed were reconvicted. Although promising, this study was based on a small sample with only five recidivists. Recently Nunes, Hermann, Maimmone and Woods (2014) examined change on the MOLEST and RAPE scales (Bumby, 1996) and found that pre-treatment, post-treatment and change scores as derived via clinical significant change methodology were unrelated to sexual recidivism.

Beggs and Grace (2011) examined a sample of 218 child molesters and found that positive changes in areas of sexual deviance, anger and hostility, anxiety, and social inadequacy were all associated with reduced recidivism rates. However, positive changes in other areas such as offence supportive attitudes and self-esteem, were not associated with reduced recidivism. Olver, Nicholaichuk and Wong (2013) examined psychometric treatment change with a sample of 267 Canadian sexual offenders. They found that positive treatment changes in areas of self esteem and anger and hostility were associated with reductions in sexual recidivism, though changes in rape myth acceptance, locus of control, social anxiety and depression were not associated with outcome. Olver, Beggs Christofferson, Grace and Wong (2013) then examined a further sample of treated sexual offenders from Canada and New Zealand. In this

larger sample, changes in sexual deviance, antisociality and responsivity were all associated with recidivism reductions.

Olver, Kingston, Nicholaichuk and Wong (in press) have conducted a further recent study examining the psychometric change on a sample of 392 treated adult male sexual offenders. For the majority of psychometric measures examined within this study, weak or non-significant relationships with recidivism were observed. However a measure of aggression and hostility (Aggression Questionnaire (AQ); Buss & Perry, 1992) did have a significant relationship with general violence and predicted general recidivism, and positive change as measured by this scale was associated with reductions in recidivism. After controlling for risk, positive changes on several AQ domains significantly predicted general recidivism reductions (though none predicted reductions in sexual recidivism). Finally, when the authors controlled for standardised residual change scores, the relationship between treatment changes and outcome improved on some measures. This provides some indication that the methodology employed to determine the predictive validity of psychometric measures may have a significant impact on research findings.

The inconsistency of the results in the studies presented could be to do with differences in the methodology employed as well as small sample sizes. Beggs and Grace (2011) compared three different methods for assessing treatment change in a sample of 218 adult male sexual offenders. They examined pre to post-treatment change on psychometrics, change on the Violence Risk Scale: Sex Offender (VRS:SO; Wong, Olver, Nicholaichuk & Gordon, 2003) tool, and post-treatment ratings on the Standard Goal Attainment Scaling for Sex Offenders (SGAS; Hogue,

1994). They found that all of these different methodologies for assessing change were positively correlated, and all were significantly related to sexual recidivism (though to differing degrees). Survival analyses revealed that all measures predicted reoffending after controlling for static risk. The overall AUC for the psychometric battery was .68, whilst for the VRS:SO it was .70 and for the SGAS it was .66.

The current trend suggests that taking into account dynamic psychological problems and deviance alongside static risk can contribute to predicting recidivism using some tools within some samples. However, the research findings seem to depend on how the psychometrics have been gathered and used, which psychometrics have been used, how the relationship between the tests and reconviction has been examined statistically, and the samples used.

Rationale for the research

Pre- and post-treatment psychometrics are used routinely to identify treatment needs or individual dynamic risk factors, to plan treatment, and to assess short-term treatment impact on the NOMS SOTPs. In order to be confident in the use of psychometrics as identifiers of treatment need, and measures of dynamic risk factors as well as measures of short term change, we need to know whether the measures and the pre and post-treatment scores are associated with recidivism. The main overarching aim of this thesis was thus to examine the discriminant and predictive validity of the psychometrics. Understanding the relationship between psychometrics and recidivism would enable risk report writers and treatment providers to provide more detailed information regarding an offender's risk of reconviction in relation to psychometric scores routinely collected, and provide more information about the

characteristics which need to change in order to reduce risk. The practical aim of this thesis was to determine the best tools to use to measure dynamic risk within sexual offenders, and as such to minimise resources and to ensure the most effective procedures are in place.

Hypotheses

The thesis intended to examine psychometric scores' relationship to recidivism. Specific and overarching hypotheses were as follows:

1) Pre-treatment and post-treatment psychometrics will demonstrate good discriminant validity.
2) Pre-treatment and post-treatment psychometrics will demonstrate good predictive validity.
3) There will be no difference in discriminant and predictive validity between pre- and post-treatment psychometrics.
4) Pre and post-treatment psychometrics will be as good as static factors in predicting reconviction.
5) Psychometric change scores will demonstrate good predictive validity.
6) A smaller psychometric battery with equivalent or better discriminant and predictive validity can be created.

Research Design

This was predominantly a quantitative investigation. As this research was not an evaluation of an intervention it does not fit neatly on the Scientific Methods Scale. There were no comparison groups being used; the research utilises only offenders

who had undergone treatment. Thus if it were to fit onto the scale it would be level 1 (within subjects correlational design). Harkins and Beech (2007) describe this design as 'within-treatment change'. However, the results are not reported as an evaluation of the interventions. Any comparisons made between groups within the research (e.g., those who appear to change on the psychometrics and those who do not appear to change), do not provide conclusions as to the effectiveness of the programme, rather they provide information regarding how psychometric scores relate to outcomes for particular groups of offenders.

Ethical Considerations

The British Psychological Society code of ethical practice (BPS, 2009) was adhered to in the research design for all empirical studies. Ethical approval was also gained from the following: the National Offender Management Service (via the National Research Committee), and the University of Birmingham's School of Psychology Ethics Committee. Copes of these approvals are provided in Appendix A.

CHAPTER TWO

AN INITIAL EXPLORATION OF THE RELATIONSHIP BETWEEN PRE AND POST-TREATMENT SELF-REPORT PSYCHOMETRICS AND RECIDIVISM

Chapter Outline

As outlined in chapter one there is inconclusive evidence as to the predictive power of self-report psychometrics. This chapter sets out to explore the relationship between individual psychometrics used on the NOMS SOTP in prison in England and Wales and recidivism outcome. Analyses are also conducted to examine the relationship between psychometrics grouped by SARN dynamic risk domain and recidivism outcome. This chapter focuses on providing descriptive statistics showing pre and post comparisons on all of the psychometric measures used on SOTP. Pre and post psychometrics were examined separately and their discriminant and predictive validity explored using a variety of different analyses; examination of change in psychometrics is examined in subsequent chapters.

Research Questions

The research questions for this chapter are:

1. Which of the pre- and post-treatment psychometrics can discriminate between recidivists and non-recidivists in a sample of treatment completers?
2. When psychometrics are grouped to form dynamic risk domain scores, in which domains do recidivists differ from non-recidivists?

3. Which of the dynamic risk domain scores are the best predictors of recidivism?
4. Do the dynamic risk domain scores add incremental validity to static risk?
5. Is there an association between the number of dysfunctional dynamic risk domains and recidivism?
6. Does the total number of dysfunctional dynamic risk domains predict recidivism?
7. Does the number of dysfunctional dynamic risk domains add incremental validity to static risk?

Method

Sample Description

A total of 3773 sexual offenders were sampled. All had committed at least one offence with a sexual element, and completed the Core or Rolling SOTP (described in Chapter One) within a prison establishment in England and Wales between 1996 and 2006. Drop outs were not included, as the intention was to use both pre- and post-treatment psychometric data. The sample had completed treatment in one of 38 establishments which offered treatment during this period of time. These establishments' security rating ranged from open to high. Table 1 below provides demographic and offence details of the sample.

Table 2.1 Sample Characteristics

Variable	*n*	*Mean (SD)*
Age at release	3773	44.66 (13.15)
	n	*%*
Sentence Length		
Under 1 year	99	2.6
1 – 2 years	364	9.6
2 – 4 years	1049	27.8
4 plus	2212	58.6
Missing	49	1.3
Ethnicity		
White	3345	88.6
Black	147	3.9
Asian	75	2.0
Other	17	0.5
Not recorded	189	5.0
Offence Type		
Adult Victim (>16)	951	25.2
Child Victim (<16)	2488	65.9
Not known	334	8.9
RM2000/s[1]		
Low	1030	27.3
Medium	1661	44.0
High	652	17.3
Very High	354	9.4
Not known	76	2.0
RM2000/v[2]		
Low	2119	56.2
Medium	1008	26.7
High	467	12.4
Very High	163	4.3
Not known	16	0.4
Reconviction		
Sexual	135	3.6
Compliance[3]	239	6.3
Violent	172	4.6
Sex or Violent	307	8.1
Any	824	21.8

The mean age of the sample at release was 44.66 (range 16-84). The mean length of follow up for the sample was 1521.95 days. Over half the sample had a sentence length of over 4 years. The majority of the sample was White. Ninety one percent of the sample had a sexual index offence. The majority of the remaining nine percent had been convicted of a violent offence. Offenders were classified into rapists or

[1] Risk Matrix 2000 Sexual Scale (Thornton et al., 2003)
[2] Risk Matrix 2000 Violence Scale (Thornton et al., 2003)
[3] Compliance offences are those which relate to the non-compliance of probation or restriction orders

child molesters based on the age of the victim of their current offence; those with an adult victim (16 years or over) were defined as 'rapists', and those with a child victim (under the age of 16) were defined as 'child molesters'. This was based on the index offence codes, therefore it is important to note that this distinction may not be entirely accurate. For example, it does not allow us to determine the age of the victim of a violent offence (with a sexual element) and it does not allow us to determine the ages of victims from previous offences. Recent research further indicates that the number of offenders who have both adult and child victims might be higher than previously assumed (Cann, Friendship & Gozna, 2007). However, it was the best available information in order to categorise the sample into 'types' of sexual offenders.

The static risk and reconviction characteristics of the sample are described below, following a description of these measures.

Materials

Static Risk: RM2000 (Thornton et al., 2003)

Static risk for the sample was produced using RM2000/s (Thornton et al., 2003), a static risk assessment tool for use with adult males who have been convicted of a sexual offence. At least one of the sexual offences must have been committed when the offender was over 16. The RM2000/s predicts sexual recidivism and is made up of seven items divided into two scoring steps. Step one comprises three items: Age of the offender on release, number of sentencing occasions for a sexual offence and number of sentencing occasions for any criminal offence. The scores assigned to each of these items are summed and translated into one of four preliminary risk categories:

Low, medium, high or very high. The second scoring step has a further four risk-raising items: any or all of the victims of sexual offending have been male, any or all of the victims of sexual offending have been strangers, the offender has never had a stable live in relationship for over two years (termed the 'single' item), and one of their sexual offences have been non-contact. If two or three of these items are present the initial risk category is raised one level (e.g., from low risk to medium). If all four of these aggravating factors are present the initial risk category is raised by two levels.

The RM2000/v, which predicts violent reoffending, is comprised of three items; age, number of sentencing occasions for a violent offence, and whether or not the offender has ever been convicted of a burglary. The items are summed and then translated into one of four risk categories. The RM2000/s and /v scales have both been shown to have good predictive validity in UK samples (Barnett, Wakeling & Howard, 2010; Craig, Beech & Browne, 2006; Grubin, 2008; Thornton et al., 2003).

Proxy RM2000/s scores were computed for the present study sample retrospectively using Police National Computer (PNC) data. Out of the seven RM2000 variables, it was possible to score five with the data available for the present sample; age, number of sexual appearances, number of criminal appearances, whether or not the offender had ever been convicted of offending against a male, and whether the offender had ever had a non-contact offence. It was not possible to score the single and stranger items with the whole sample, as this information was not available or complete for all offenders from the treatment databases or the PNC data. RM2000/s scores were calculated for those in the sample for whom the single and stranger items were the only two missing items. This conforms with other research studies (e.g., Barnett et

al., 2010; Langton, Barbaree, Hanson, Harkins & Peacock, 2007). Since this revised RM2000/s scoring was likely to result in consistent underscoring, those who were scored as having one or both of the two available aggravating factors (male victim and non-contact) were raised a risk category. The similarities between these risk scores and a subsample for whom full RM2000/s scores were available (N = 1924, 51% of sample) were examined. Approximately 64% of these were in the same risk categories using both the full and revised scoring procedures, 9% were rated as a higher risk category using the revised scoring procedure, and 27% were rated as a lower risk category using the revised scoring procedure. Thus the revised scoring procedure resulted in a greater level of underscoring. According to current guidance on using RM2000 with internet only offenders (Thornton, 2010), the non-contact and stranger items should not be scored. The present study adheres to these guidelines; it was not possible to score the male item for the sample (because the offence codes relating to indecent images of children do not indicate the gender of the victims in the images), and since the non-contact item should not be scored, all of these offenders remained at their step one risk category.

Forty four percent of the present sample were medium risk on the /s scale, while 9.4 percent were very high risk. Regarding the /v scale, the majority of the sample were low risk (56.2%), while only 4.3 percent were very high risk. Regarding individual RM2000/s items, the average number of sexual appearances was 1.51, and the average number of criminal appearances was 4.22. Twenty-seven percent of the sample had a male victim, and sixteen percent of the sample had committed a non-contact sexual offence.

Reconviction Data

Reconviction data was sourced from the Police National Computer (PNC) for all offenders who had been released prior to April 2008. This data included the date of the first reconviction, and the type of offence the reconviction referred to. Different reconviction variables were computed based on the type of offence for which offenders were reconvicted for (e.g., sexual, sexual and/or violence, violence, compliance, general).

As can be seen in Table 2.1, 3.6 percent of the sample had been reconvicted of a sexual offence (N = 135), 6.3 percent had been reconvicted of a compliance offence (N = 239), 4.6 percent had been reconvicted of a violent offence (N = 172), 8.1 percent had been reconvicted of a sexual or a violent offence (N = 307), and 21.8 percent had been reconvicted for any offence (N = 824). These rates relate to any conviction after release from prison.

The outcome measure of interest within the present study is sexual and violent recidivism. This outcome was selected, as opposed to sexual recidivism alone, due to the higher base rates of sexual and violent recidivism within the present sample, and taking into account existing evidence that sexual offenders will often re-offend with a non-sexual offence (Hanson & Bussière, 1998). Numerous studies have found that those convicted of sexual offences present an equal and sometimes greater risk of committing nonsexual violent reoffences, as they do sexual reoffences (e.g., Thornton & Travers, 1991; Grubin, 2008; Barnett, Wakeling & Howard, 2010).

Psychometric Measures

A battery of psychometric measures are administered before and after every SOTP, in order to help determine treatment need prior to commencing treatment, and to help measure progress in treatment. In 2003 a number of additional measures were added to the psychometric battery used within SOTP. Therefore, all participants have scores for most of the measures (described henceforth as the Old Battery – OB), whilst only a proportion of participants have scores for those measures added in 2003 (described henceforth as the New Battery – NB). All psychometric measures have been examined within this chapter even those with poorer reliability and validity, since it was felt useful to include all of the gathered and available data. Empirical analyses for subsequent chapters take a different stance.

The Main Psychometric Measures – Old Battery

The measures in table 2.2 all formed part of the Old Battery. Full details of these measures can be found in Appendix B.

The New Psychometric Measures – New Battery

The psychometric measures in table 2.3 were added to the pre and post assessment battery in 2003. Again full details of these can be found in Appendix B.

Table 2.2 Old Battery Psychometrics

Measure	*Reference(s)*	*Construct of interest*	*Number of items*	*Response Format*	*Scale score range*	*Internal Consistency (α)*	*Test retest Reliability (r)*
Entitlement to sex	Hanson, Gizzarelli & Scott (1994)	Beliefs about entitlement to sex	9	5 point likert scale	0-36	0.84	0.67
Impulsivity Scale	Eysenck & Eysenck (1978)	Tendency to act impulsively	13	Yes/No Dichotomous	0-26	0.84	0.79
The Interpersonal Reactivity Index	Davis (1980)	Cognitive and emotional components of empathy	28	5 point likert scale	0-28 for each subscale	Fantasy 0.77 Empathic concern 0.72 Perspective taking 0.72 Personal distress 0.74	Fantasy 0.77 Empathic concern 0.79 Perspective taking 0.81 Personal distress 0.74
Locus of Control Scale	Levenson (1974)	Tendency to equate experiences to internal or external factors	18	5 point likert scale	0-72	0.79	0.87
Openness to Men and Women	Underhill, Wakeling, Mann & Webster (2008)	Intimacy with adult males and females	18	5 point likert scale	0-36 for each subscale	Openness to men 0.85 Openness to women 0.86	Openness to men 0.86 Openness to women 0.81
Relapse Prevention Questionnaire	Beckett, Fisher, Mann & Thornton (1997)	Recognition of lapse cues, possession of coping skills and strategies	18	3 point scale (0-2)	0-34	Recognition of lapse cues 0.85 Coping skills and strategies 0.80 Total scale 0.85	
Revised Dissipation Rumination Scale	Caprara (1986) Wakeling & Barnett (2011)	Tendency to bear grudges and ruminate	15	Yes/No dichotomous	0-30	0.78	0.64
Self Esteem Scale	Webster, Mann, Thornton & Wakeling (2007)	General self esteem	8	Yes/No dichotomous	0-16	0.84	0.90
Sex Offence Attitudes Questionnaire	Hogue (1994)	Levels of cognitive denial, distortion and minimisation	50	5 point likert scale	33-132	Denial of repetition 0.74 Denial of premeditation 0.78 Denial of harm 0.85 Denial of offence 0.49	Denial of repetition 0.85 Denial of premeditation 0.86 Denial of harm 0.44 Denial of offence 0.83

						Denial of responsibility 0.84 Denial of control 0.73 Total scale 0.86	Denial of responsibility 0.82 Denial of control 0.69 Total scale 0.82
Sex with Children is Justifiable	Mann, Webster, Wakeling & Marshall (2007)	Extent to which an individual believes that children enjoy sexual contact with adults	18	5 point likert scale	0-72	0.94	0.93
The University of California (UCLA) Loneliness Scale	Russell, Peplan & Cutrona (1980)	Loneliness	19	4 point likert scale	19-76	0.95	0.79
Women are deceitful Scale	NOMS (unpublished)	Beliefs about women	5	5 point likert scale	0-20	0.79	0.81

Table 2.3 New Battery Psychometrics

Measure	*Reference(s)*	*Construct of interest*	*Number of items*	*Response Format*	*Scale score range*	*Internal Consistency (α)*	*Test retest Reliability (r)*
Children and Sex Questionnaire	Beckett (1987)	Attitudes, feelings and thoughts about children and sex	87	5 point likert scale	0-60 for each subscale		Cognitive Distortions 0.77 Emotional Congruence 0.63
The Emotion Control Questionnaire	Roger & Najarian (1989)	Emotional response style	56	True/False Dichotomous	0-28 for each subscale	Rehearsal 0.86 Emotional Inhibition 0.77 Aggression Control 0.81 Benign Control 0.79	Rehearsal 0.80 Emotional Inhibition 0.79 Aggression Control 0.73 Benign Control 0.92
The Hypermasculinity Inventory	Mosher & Sirken (1984)	Macho personality constellation	30	Forced choice Dichotomous	0-60	Calloused sex 0.79 Violence as manly 0.79 Danger as exciting 0.71 Total scale 0.89	
The Multiphasic Sex Inventory	Nichols & Molinder (1984)	Beliefs, attitudes and behaviours surrounding sex	300	True/False Dichotomous	0-300	Sexual obsessions 0.89	Sexual obsessions 0.88
Relationship Style Questionnaire	Dutton, Saunders, Starzomski & Bartholomew (1994)	Attachment patterns	30	5 point likert scale	19-95	Secure 0.41 Dismissing 0.71	
Social Problem Solving Inventory – Revised	D'Zurilla, Nezu & Maydeu-Olivares (2002)	Ability to resolve every day life problems	52	5 point likert scale	0-208	Positive problem orientation 0.73 Negative problem orientation 0.92 Rational problem solving 0.95 Impulsivity/Carelessness Style 0.91 Avoidance Style 0.92	Total scale 0.94
Paulhus Scale	Paulhus (1984)	Social desirable responding	40	7 point likert scale	0-40	Self-deception 0.72 Impression management 0.84 Total scale 0.86	

Procedure

Pre- and post-treatment psychometric scores are held centrally by staff at NOMS Interventions Services for all offenders who have undertaken Sex Offender Treatment in a UK prison. Offenders provide consent for their assessment data to be used for research purposes when they embark on a treatment programme. Those participants who did not consent for their data to be used were removed. Reconviction data was sourced for the treatment sample.

Prior to analysis the psychometrics were mapped onto the relevant SARN dynamic risk domains, as described in chapter one. Please refer to Table 2.4 below for a list of which psychometrics intend to measure which dynamic risk factor. Three of the subscales do not map onto any of the dynamic risk domains: minimisation, impression management, and self deception enhancement.

Table 2.4 Psychometrics by SARN Dynamic Risk Domain

Dynamic Risk Domain	Psychometric Measure - subscale	Old (OB) or New Battery (NB)
Sexual Interests	MSI Sexual Obsessions	NB
	MSI Paraphilias	NB
Offence Supportive Attitudes	Women are Deceitful	OB
	Entitlement to Sex	OB
	Sex with Children	OB
	Children and Sex Cognitive Distortions	NB
	Danger as Exciting	NB
	Violence is Manly	NB
	Callous Attitudes	NB
Socio-Affective Functioning	Ruminations	OB
	Emotional Loneliness	OB
	Self Esteem	OB
	Empathic Concern	OB
	Perspective Taking	OB
	Locus of Control	OB
	Openness to Men	OB
	Openness to Women	OB
	Emotional Congruence with Children	NB
	Secure Relationship Style	NB
	Fearful Relationship Style	NB
	Preoccupied Relationship Style	NB
	Dismissing Relationship Style	NB
Self Management	Impulsivity	OB
	Personal Distress	OB
	Fantasy	OB
	Relapse Prevention Coping	OB
	Relapse Prevention Recognition	OB
	Rehearsal	NB
	Emotional Inhibition	NB
	Aggression Control	NB
	Benign Control	NB
	Positive Problem Solving	NB
	Negative Problem Solving	NB
	Rational Problem Solving	NB
	Impulsive Problem Solving	NB
	Avoidant Problem Solving	NB
	Problem Solving Total	NB
Measures not specifically targeting dynamic risk factors	Minimisations	OB
	Impression Management	NB
	Self Deception Enhancement	NB

Data Analytic Plan

A variety of statistical analyses were employed to answer the research questions for this chapter. The relationship between static risk and recidivism was initially explored using t-test analyses, and Receiver operating characteristic analysis (ROC). ROC analysis was used to establish the predictive accuracy of the RM2000/s for different types of reoffending. The ROC charts the proportion of people correctly classified as recidivists against the proportion of people falsely classified as recidivists at different decision thresholds. The resulting statistic called the Area under the Curve statistic (AUC) represents the accuracy of the scale's classification of recidivists. There is ongoing debate about how to interpret AUCs; the present study follows the interpretation suggested by Douglas, Epstein and Poythress (2008), where comparisons to Cohen's *d* suggest that an AUC value of 0.71 represents a large effect and an AUC value of 0.64 represents a moderate effect.

Exploratory univariate analyses were then used to explore the differences between recidivists and non-recidivists on the individual self-report psychometric measure scores.

1) Firstly, t-tests were used to examine any differences between recidivists and non-recidivists on raw pre- and post-treatment psychometric scores (discriminant validity).

2) Next exploratory Cox regression analyses were performed using SARN risk domain scores generated from the psychometric scores contributing to each risk

domain (predictive validity). Incremental validity of these risk domain scores was also examined by entering them into regression analyses with static risk. Cox regressions were used rather than logistic regression, due to the fact that censored time-to-event data was used (Harrell, Lee & Mark, 1996). For all regression analyses, the number of variables which can be entered into the models is based on Harrell et al.'s (1984) formula which states that m/10 number of predictors can be used, where m is the number of people in the less frequent outcome category (here the reconviction group). When using sexual and/or violent recidivism as the outcome variable, 30 variables can be entered into the model (307 offenders were reconvicted for a sexual or violent offence; 307/10 = 30.7). This number varies for each analysis dependent on the number of offenders in the reconviction group for whom the psychometric data is available.

The dynamic risk domain scores were calculated by standardising each of the psychometric measures using the mean and standard deviation of the entire sample. Once standardised, scores greater than 0 were considered dysfunctional, as this represents scoring lower than the average sexual offender. For those psychometrics on which high scores denote greater functionality than low scores, the standardised scores were initially multiplied by -1, so that greater than 0 was dysfunctional across all psychometrics. Domain scores were then created by summing the standardised scores of all of the psychometrics relating to each of the four dynamic risk domains. Please refer back to Table 2.4 for information regarding which psychometrics relate to which dynamic risk domains. The decision as to which psychometrics relate to which dynamic risk domain was made theoretically, based on the constructs the psychometrics intend to measure.

It should be noted that those psychometrics not specifically measuring a particular dynamic risk domain (e.g., minimisation) were not included in these analyses. Missing psychometric scores were given the average standardised score for that psychometric. If 10% of scores within a domain were missing, a total domain score was not computed. If less than 10% of scores with a domain were missing, these missing scores were imputed with the average standardised scores. The domain scores were then standardised themselves; if the score was over 0 the domain was counted as dysfunctional. The possibility of creating dynamic risk domain scores using factor analytical techniques was examined; there were no differences in the outcome using this methodology, therefore the decision was taken to proceed using the original dynamic risk domain scores, which weight all psychometrics within a domain equally.

3) Finally a Psychological Deviance Index (PDI) was produced, which is the number of dysfunctional domains for each individual, which can range from 0 to 4, with high scores indicative of greater dysfunction. The PDI can be used as a measure of psychological deviance. T-tests, chi squares and Cox regression models were then performed on these dynamic risk domain scores and overall PDI scores to answer the research questions outlined previously.

Results

RM2000/s and Recidivism

A total of 27.9% of the sample were classified as low risk, 44.9% as medium risk, 17.6% as high risk, and 9.6% as very high risk. The mean RM2000/s scores were significantly higher for those men who sexually reoffended (M= 2.00, SD= 1.03) than for those who did not (M= 1.05, SD= 0.89) (t (140.55) = -10.57, $p < .001$). The mean RM2000/s scores were also significantly higher for those men who sexually or violently reoffended (M= 1.65, SD= 0.95) than for those who did not (M= 1.04, SD= 0.89) (t (345.88) = -10.62, $p < .001$). The higher risk categories were associated with higher rates of sexual, sexual and violent, and general recidivism. The AUCs for sexual, sexual and violent and general recidivism respectively were .75, .67 and .62. Thus the predictive validity for the RM2000/s was moderate for sexual and violent, and general recidivism, and good for sexual recidivism.

Self-report Psychometric Data – Exploratory Univariate Analyses

RQ1: Which of the pre- and post-treatment psychometrics can discriminate between recidivists and non-recidivists?

The means and standard deviations for all the self-report psychometric measures both pre- and post-treatment are shown in Table 2.5. Sexual and violent recidivists and non-recidivists' scores are both shown. Data for the NB self-report measures has only been gathered since 2003, therefore sample size varies for different measures. A series of t-tests were conducted to examine differences between recidivists and non-recidivists on this univariate psychometric data both pre and post-treatment. Due to

the fact that a large number of t-tests were performed, the Bonferroni adjustment was used to adjust the p- value. Using this adjusted *p*-value, at the pre-treatment stage there were significant differences on the following measures between groups: entitlement to sex, ruminations, self esteem, empathic concern, perspective taking, locus of control, and impulsivity. On one measure however, recidivists scored worse than the non-recidivists; the Impression Management subscale. That is, the recidivists had significantly lower levels of impression management than non-recidivists. At the post-treatment stage there were significant differences on fewer measures between groups (with the Bonferroni adjustment): ruminations, empathic concern, perspective taking, impulsivity, and impression management. Recidivists scored significantly worse than non recidivists on all of these measures, apart from impression management, on which the non-recidivists scored worse than the recidivists.

For the child molester specific measures (Children and Sex Cognitive Distortions, Sex with Children, and Emotional Congruence with Children) means and standard deviations are also reported for the child molesters only in the sample. To examine whether differences between groups on these child molester scales would be significant when comparing the child molester recidivists with the child molester non recidivists, further t-tests were computed with this reduced sample. No significant differences were found between child molester recidivists and non-recidivists pre- or post-treatment.

Table 2.5 Pre and Post Psychometric Measure Scores

Psychometric	Pre-treatment Score (SD)		Post-treatment Score (SD)	
	Sexual and violent Recidivists	Non-sexual and non-violent Recidivists and Non Recidivists	Sexual and violent Recidivists	Non-sexual and non-violent Recidivists and Non Recidivists
SEXUAL INTERESTS				
MSI Sexual Obsessions	4.17 (4.83)*	2.29 (3.06)	2.39 (2.89)	1.57 (2.58)
MSI Paraphilias	3.66 (4.18)*	2.33 (3.89)	2.82 (3.51)	1.89 (4.07)
OFFENCE SUPPORTIVE ATTITUDES				
Children and Sex Cognitive Distortions	10.17 (10.11)	10.83 (9.36)	5.29 (6.93)	5.18 (6.30)
Child molesters only	11.52 (11.36)	11.85 (9.57)	5.60 (8.38)	5.58 (6.59)
Sex with Children	6.85 (9.88)	6.82 (9.02)	2.34 (6.01)	2.18 (5.11)
Child molesters only	9.35 (11.21)	7.94 (9.54)	2.87 (7.11)	2.43 (5.46)
Danger as Exciting	4.05 (3.37)**	2.39 (2.83)	2.59 (2.84)*	1.82 (2.25)
Violence is Manly	2.76 (3.23)	1.88 (2.93)	1.26 (2.01)	1.11 (2.20)
Callous Attitudes	1.41 (1.38)	1.32 (1.54)	1.09 (1.38)	1.08 (1.48)
Women are Deceitful	8.10 (3.92)*	7.58 (3.94)	5.19 (3.61)	4.82 (3.65)
Entitlement to Sex	10.49 (4.49)***	9.38 (4.23)	7.45 (4.13)*	6.87 (4.06)
SOCIO-AFFECTIVE FUNCTIONING				
Ruminations	11.80 (7.13)***	8.96 (6.40)	8.36 (6.51)***	6.38 (6.04)

Emotional Loneliness	45.66 (13.07)**	43.03 (13.43)	-	-
Self Esteem	7.56 (5.01)***	8.77 (5.21)	10.38 (5.12)**	11.25 (4.84)
Empathic Concern	18.49 (4.69)***	20.24 (4.15)	18.88 (4.56)***	20.17 (4.47)
Perspective Taking	15.77 (5.36)***	17.77 (5.15)	17.69 (5.36)***	19.54 (5.18)
Locus of Control	45.30 (8.62)***	47.66 (8.57)	51.57 (9.01)**	53.25 (8.69)
Openness to Men	21.12 (6.21)**	22.26 (5.76)	23.17 (5.80)*	24.14 (5.72)
Openness to Women	23.18 (6.01)	23.38 (5.68)	24.21 (5.26)	24.77 (5.43)
Emotional Congruence with Children	14.70 (10.33)	14.77 (10.64)	10.33 (9.20)	11.28 (9.87)
Child molesters only	15.52 (7.98)	14.89 (10.38)	9.61 (7.81)	10.94 (9.30)
Secure Relationship Style	13.66 (3.53)**	15.74 (3.21)	15.81 (4.15)	17.05 (3.32)
Fearful Relationship Style	15.33 (3.98)	14.46 (3.92)	14.33 (3.66)*	12.86 (3.44)
Preoccupied Relationship Style	10.75 (2.72)	10.84 (3.04)	10.47 (2.99)	10.03 (2.73)
Dismissing Relationship Style	16.78 (3.42)	15.93 (4.00)	16.64 (3.64)*	15.22 (3.70)
SELF MANAGEMENT				
Impulsivity	13.94 (6.83)***	10.37 (7.22)	10.15 (7.52)***	7.40 (7.12)
Personal Distress	11.68 (5.15)*	10.93 (5.20)	10.09 (5.09)*	9.36 (5.15)
Fantasy	12.96 (5.64)*	12.25 (5.79)	12.39 (6.07)	11.83 (5.98)
Relapse Prevention Coping	7.13 (3.27)	7.02 (3.49)	11.32 (3.00)	11.41 (3.23)
Relapse Prevention Recognition	8.47 (4.58)**	7.45 (4.60)	13.14 (3.87)	12.63 (4.07)
Rehearsal	9.85 (6.77)*	7.27 (5.68)	5.11 (4.01)	5.08 (4.13)
Emotional Inhibition	15.73 (6.36)	14.21 (6.09)	11.47 (6.28)	11.06 (5.83)

Aggression Control	17.60 (6.87)*	20.51 (5.93)	20.03 (5.80)	21.59 (4.95)
Benign Control	14.61 (8.29)*	17.74 (7.01)	20.00 (6.70)	21.65 (5.90)
Positive Problem Solving	12.05 (4.45)	12.15 (3.81)	14.24 (3.60)	13.84 (3.63)
Negative Problem Solving	14.33 (9.72)	11.80 (8.67)	8.49 (7.84)	7.04 (6.93)
Rational Problem Solving	41.78 (17.92)	44.97 (16.01)	52.41 (17.08)	54.13 (15.63)
Impulsive Problem Solving	14.70 (8.90)*	11.70 (8.21)	9.32 (7.58)	6.91 (6.97)
Avoidant Problem Solving	10.63 (6.70)	8.84 (5.50)	7.08 (4.40)	6.10 (4.60)
Problem Solving Total	12.08 (3.71)	13.06 (3.27)	14.68 (3.11)	15.21 (2.92)
MEASURES NOT RELATED TO DYNAMIC RISK FACTOR DOMAINS				
Minimisations	81.87 (17.55)*	84.50 (17.15)	64.26 (14.96)**	66.83 (14.90)
Impression Management	4.77 (3.12)***	6.73 (4.17)	4.77 (3.12)***	6.73 (4.17)
Self Deception	5.44 (2.98)	5.70 (3.44)	5.44 (2.98)	5.70 (3.44)

* significant at the .05 level. ** significant at the .01 level. *** significant at the .001 level. Before Bonferroni adjustments were made.

Multivariate Analyses: Dynamic Risk Domain scores

RQ2: When psychometrics are grouped to form dynamic risk domain scores, in which domains do recidivists differ from non-recidivists?

T-test analyses found that recidivists scored significantly higher (more problematic) on the following dynamic risk domain scores using the old psychometric battery pre-treatment: offence supportive attitudes domain ($t(323.78) = -2.64$, $p < .01$), socio-affective functioning domain ($t(332.94) = -7.61$, $p < .001$), and the self management domain ($t(335.72) = -5.88$, $p < .001$); and the following domain scores using the new psychometric battery: sexual interests domain ($t(42.59) = -2.58$, $p < .05$), offence supportive attitudes domain ($t(46.46) = -2.55$, $p < .05$), socio-affective functioning domain ($t(42.40) = -3.69$, $p < .01$), and the self management domain ($t(41.50) = -3.84$, $p < .001$). Post-treatment recidivists scored significantly higher (more problematic) on the following domain scores using the old psychometric battery: socio-affective functioning domain ($t(295.39) = -5.42$, $p < .001$), and the self management domain ($t(302.65) = -4.28$, $p < .001$); and the socio-affective functioning domain only using the new psychometric battery ($t(37.23) = -2.26$, $p < .05$).

RQ3: Which of the dynamic risk domain scores are the best predictors of recidivism?

Cox Regression analyses were next performed to examine the ability of the dynamic risk factor domain scores to predict recidivism. Four separate regression analyses were performed using time at risk, and sexual and violent recidivism as the outcome (pre and post-treatment scores for OB and NB separately). When examining the pre-treatment domain scores using the new psychometric battery, the model was significant ($-2LL = 454.24$, $p < .01$), but none of the individual domain scores were significant predictors of outcome. When examining the pre-treatment domain scores using the old psychometric battery, the model was significant ($-2LL = 4141.59$, $p < .001$). The socio-affective functioning domain was the only significant

predictor of outcome; greater problems in social affective functioning were associated with greater recidivism.

Post-treatment, the model for the domain scores using the new psychometric battery was significant (*-2LL* = 391.78, $p < .05$). Again the socio-affective functioning domain was the only significant predictor, with greater problems in the socio affective functioning area being associated with greater sexual and/or violent recidivism. Lastly, when examining the post-treatment domain scores using the old psychometric battery, the model was significant (*-2LL* = 3804.33, $p < .001$). Again the socio-affective functioning domain was the only significant predictor of outcome, with greater problems in this area being associated with greater sexual and/or violent recidivism.

Table 2.6 Cox Regression Analyses for Pre- and Post-treatment Risk Domain Scores by Battery Type

	B	SE	Wald	Sig.	Exp(B)
Model 1: NB Pre-Treatment					
Domain 1 NB Pre Total Score	.109	.061	3.163	.075	1.115
Domain 2 NB Pre Total Score	-.006	.042	.020	.887	.994
Domain 3 NB Pre Total Score	.036	.033	1.177	.278	1.036
Domain 4 NB Pre Total Score	.091	.058	2.446	.118	1.095
Model 2: OB Pre-Treatment					
Domain 2 OB Pre Total Score	-.013	.027	.238	.626	.987
Domain 3 OB Pre Total Score	.069	.014	22.881	.000	1.071
Domain 4 OB Pre Total Score	.055	.033	2.771	.096	1.057
Model 3: NB Post-Treatment					
Domain 1 NB Post Total Score	.070	.076	.843	.359	1.072
Domain 2 NB Post Total Score	-.069	-054	1.650	.199	.933
Domain 3 NB Post Total Score	.072	.037	3.862	.049	1.075
Domain 4 NB Post Total Score	.033	.059	.318	.573	1.034
Model 4: OB Post-Treatment					
Domain 2 OB Post Total Score	-.028	.030	.882	.348	.972
Domain 3 OB Post Total Score	.049	.017	8.630	.003	1.050
Domain 4 OB Post Total Score	.034	.034	.980	.322	1.035

Domain 1 = Sexual Interests, Domain 2 = Attitudes, Domain 3 = Socio-affective functioning, Domain 4 = Self Management

RQ4: Do dynamic risk domain scores add incremental validity to static risk?

To examine whether dynamic risk domain scores provide increased predictive validity for sexual and violent recidivism while controlling for static risk level, a series of hierarchical Cox regressions were conducted. For these analyses, the RM2000/s risk categories (Categorical variable) were entered into the model at the first step, and the domain scores

were entered individually at the second step in a forward stepwise model. Four separate analyses were performed but all results are shown in table 2.7.

The first analysis performed examined the incremental predictive validity of the pre-treatment NB domain scores. At step 1, the RM2000 variables were entered, and then at step 2 the four new battery dynamic risk domain scores were entered using a stepwise method. The final model was significant (*-2LL* = 413.39, $p < .001$), but the only significant variable in the model was the pre-treatment self management domain score; none of the RM2000 categories or other psychometric domain scores individually were predictive of outcome. A similar finding emerged when examining the incremental validity of the post-treatment NB domain scores. The final model was significant (*-2LL* = 361.53, $p < .05$), but the only significant variable in the model was the post-treatment socio-affective functioning domain score.

The next regression examined the incremental predictive validity of the pre-treatment OB dynamic risk domain scores. The overall model was significant (*-2LL* = 3909.883, $p < .001$). The socio-affective functioning domain 3 was a significant predictor of sexual and violent recidivism, and added to the predictive validity of the RM2000/s risk categories alone (added predictive validity, $\chi^2 = 21.04$, $p < .00$). This analysis was repeated using the sample for whom we had full RM2000 scores, as the additive power of the relationships dynamic risk domain could have been due to the lack of the relationship item in the static risk scores used in the current research. However the model was the same when using full RM2000/s scores; that is pre-treatment socio-affective functioning domain scores contributed incrementally to the prediction of reconviction over and above static risk.

Table 2.7 Summary of Regression Analysis Examining Incremental Validity of Pre- and Post-Treatment Dynamic Risk Domain Scores by Battery Type

	B	SE	Wald	Sig.	Exp(B)
Model 1: NB Pre-Treatment					
Step 1 (Enter method)					
RM2000 Low Risk Category			4.239	.237	
RM2000 Medium Risk Category	.359	.488	.542	.462	1.432
RM2000 High Risk Category	.891	.527	2.859	.091	2.439
RM2000 Very High Risk Category	1.025	.646	2.518	.113	2.787
Step 2 (Stepwise method)					
RM2000 Low Risk Category			2.501	.475	
RM2000 Medium Risk Category	.167	.493	.115	.734	1.182
RM2000 High Risk Category	.707	.530	1.784	.182	2.028
RM2000 Very High Risk Category	.563	.662	.723	.395	1.756
Self Management Domain NB Pre Total Score	1.286	.364	12.496	.000	3.618
Model 2: NB Post-Treatment					
Step 1 (Enter method)					
RM2000 Low Risk Category			4.501	.212	
RM2000 Medium Risk Category	.610	.521	1.372	.241	1.841
RM2000 High Risk Category	1.077	.558	3.726	.054	2.935
RM2000 Very High Risk Category	-.217	1.096	.039	.843	.805
Step 2 (Stepwise method)					
RM2000 Low Risk Category			4.269	.234	
RM2000 Medium Risk Category	.643	.521	1.521	.218	1.902
RM2000 High Risk Category	1.050	.559	3.536	.060	2.859
RM2000 Very High Risk Category	-.209	1.096	.036	.849	.812
Socio-affective functioning NB Post Total Score	.772	.314	6.044	.014	2.165
Model 3: OB Pre-Treatment					
Step 1 (Enter method)					
RM2000 Low Risk Category			107.469	.000	
RM2000 Medium Risk Category	1.237	.229	29.192	.000	3.446

RM2000 High Risk Category	1.562	.244	40.928	.000	4.770
RM2000 Very High Risk Category	2.346	.242	94.105	.000	10.443
Step 2 (Stepwise method)					
RM2000 Low Risk Category			88.693	.000	
RM2000 Medium Risk Category	1.191	.229	26.971	.000	3.289
RM2000 High Risk Category	1.465	.245	35.737	.000	4.329
RM2000 Very High Risk Category	2.185	.244	79.901	.000	8.887
Socio-affective functioning Domain OB Pre Total Score	.454	.098	21.377	.000	1.574
Model 4: OB Post-Treatment					
Step 1 (Enter method)					
RM2000 Low Risk Category			88.551	.000	
RM2000 Medium Risk Category	1.194	.230	27.074	.000	3.301
RM2000 High Risk Category	1.493	.247	36.531	.000	4.452
RM2000 Very High Risk Category	2.222	.247	80.804	.000	9.228
Step 2 (Stepwise method)					
RM2000 Low Risk Category			79.755	.000	
RM2000 Medium Risk Category	1.191	.230	26.911	.000	3.289
RM2000 High Risk Category	1.454	.248	34.486	.000	4.278
RM2000 Very High Risk Category	2.147	.249	74.520	.000	8.558
Socio-affective functioning Domain OB Post Total Score	.262	.089	8.556	.003	1.299

The next analyses examined the post-treatment OB psychometric domain scores. The final model was significant (*-2LL* = 3641.34, $p < .001$), and the same finding emerged; that is the socio-affective functioning domain post-treatment scores significantly predicted outcome in addition to RM2000 static risk categories (added predictive validity, $\chi^2 = 8.38$, $p < .01$), as shown in table 2.7.

RQ5: Is there an association between the number of dysfunctional dynamic risk factor domains (PDI) and recidivism?

Chi square analyses were computed to examine the association between the PDI groups and recidivism outcome. The proportion of offenders who were reconvicted of another sexual or violent offence increases with each PDI group, when using both the OB (pre- and post-treatment) and the NB (pre-treatment). The associations by PDI groups and recidivism outcome were significant when examining the OB psychometrics pre-treatment (χ^2 (3,1) = 41.24, p <. 001), and post-treatment (χ^2 (3,1) = 18.86, p <. 001), and the NB battery pre-treatment (χ^2 (4,1) = 24.22, p <. 001), but not post-treatment (χ^2 (4,1) = 8.80, p > .05). The table below shows the percentage of sexual and violent recidivism within each PDI rating.

Table 2.8 Sexual and/or Violent Recidivism by Number of Dysfunctional Dynamic Risk Factor Domains

	PDI Rating				
	0	1	2	3	4
Pre-Treatment PDI OB	3.4 %	7.6 %	9.4 %	11.7 %	N/A
Post-Treatment PDI OB	5.0 %	7.1 %	9.4 %	10.0 %	N/A
Pre-Treatment PDI NB	1.0 %	1.3 %	2.1 %	6.7 %	7.1 %
Post-Treatment PDI NB	2.0 %	1.3 %	3.4 %	3.5 %	6.3 %

RQ6: Does the total number of dysfunctional dynamic risk domains predict recidivism?

PDI scores were entered into two Cox Regression analyses, one in which the NB psychometrics pre and post PDI scores were entered, and the second in which the OB pre and post PDI scores were entered. Both sets of results are shown in Table 2.9. The first model was significant (-*2LL* = 345.78, p < .001), and the pre PDI score significantly predicting

recidivism outcome. The second model was also significant (*-2LL* = 3559.56, $p < .001$); the pre PDI score again significantly predicted recidivism outcome.

Table 2.9 Summary of Regression Analyses of PDI Scores and Sexual and/or Violent Recidivism by Battery Type

	B	SE	Wald	Sig.	Exp(B)
Model 1: New Battery					
Pre NB PDI Score	.669	.181	13.688	.000	1.953
Post NB PDI Score	.071	.156	.210	.647	1.074
Model 2: Old Battery					
Pre OB PDI Score	.303	.073	17.038	.000	1.354
Post OB PDI Score	-.016	.070	.056	.813	.984

RQ7: Does the number of dysfunctional dynamic risk domains add incremental validity to static risk?

A final two Cox regressions were performed to examine the ability of the PDI scores (new and old battery) to predict outcome along with RM2000/s risk categories (see Table 2.10). The first of these examined the incremental validity of the NB PDI scores. At step 1, the RM2000 variables were entered, and then at step 2 the two NB PDI scores (pre and post-treatment) were entered using a stepwise method. The final model was significant (*-2LL* = 316.62, $p < .001$), but the only significant variable in the model was the pre-treatment PDI score and adding this significantly improved the model ($\chi^2 = 22.02$, $p < .001$); none of the RM2000 categories or the post-treatment PDI score individually were predictive of outcome. The same was performed using the PDI scores from the OB battery. The overall model was significant (*-2LL* = 3402.90, $p < .001$). The pre-treatment OB PDI score was a significant

predictor of sexual and violent recidivism, and added to the predictive validity of the RM2000/s risk categories alone (added predictive validity, χ^2 =9.87, p < .01).

Table 2.10 Summary of Regression Analysis Examining Incremental Validity of PDI Scores by Battery Type

Variable	B	SE	Wald	Sig.	Exp(B)
Model 1: New Battery					
Step 1 (Enter method)					
RM2000 Low Risk Category			3.405	.333	
RM2000 Medium Risk Category	.517	.526	.965	.326	1.677
RM2000 High Risk Category	.968	.570	2.883	.090	2.633
RM2000 Very High Risk Category	-.150	1.096	.019	.891	.861
Step 2 (Stepwise method)					
RM2000 Low Risk Category			2.573	.462	
RM2000 Medium Risk Category	.436	.526	.686	.408	1.546
RM2000 High Risk Category	.718	.572	1.580	.209	2.051
RM2000 Very High Risk Category	-.573	1.098	.273	.602	.564
Pre-Treatment New Battery PDI Score	.709	.168	17.801	.000	2.032
Model 2: Old Battery					
Step 1 (Enter method)					
RM2000 Low Risk Category			87.201	.000	
RM2000 Medium Risk Category	1.239	.240	26.745	.000	3.451
RM2000 High Risk Category	1.551	.257	36.363	.000	4.717
RM2000 Very High Risk Category	2.290	.257	79.129	.000	9.871
Step 2 (Stepwise method)					
RM2000 Low Risk Category			72.951	.000	
RM2000 Medium Risk Category	1.214	.240	25.652	.000	3.366
RM2000 High Risk Category	1.484	.258	33.079	.000	4.412
RM2000 Very High Risk Category	2.155	.261	68.305	.000	8.628
Pre-Treatment Old Battery PDI Score	.198	.064	9.680	.002	1.219

Chapter Discussion

The present study found that a number of dynamic variables can be useful in discriminating recidivists from non-recidivists, and in predicting recidivism. A number of the pre- and post-treatment psychometrics, and dynamic risk factor domain scores were able to discriminate recidivists from non-recidivists and predict recidivism. This preliminary analysis therefore suggests that psychometric variables can be useful indicators of recidivism, and that taking into account dynamic psychological problems alongside static risk can advance our recidivism prediction.

Exploratory analyses found that proxy RM2000 scores were useful in discriminating recidivists from non-recidivists; those who reoffended had significantly higher RM2000 scores than those who did not reoffend. Thus, even though it was not possible to create full RM2000 scores for the whole sample, the results suggest that the proxy scores were useful predictors of recidivism.

When examining the individual pre- and post-treatment psychometric scores, some were able to distinguish recidivists from non-recidivists. Specifically at the pre-treatment stage, recidivists scored significantly worse on the following psychometrics than non-recidivists: entitlement to sex, ruminations, self-esteem, empathic concern, perspective taking, locus of control and impulsivity. At the post-treatment stage recidivists scored significantly worse on fewer psychometrics. However, it was four of the seven measures which significantly discriminated recidivists at the pre-treatment stage which remained useful in discriminating offenders at the post-treatment stage: ruminations, empathic concern, perspective taking and impulsivity. These particular psychometric measures therefore appear to be good indicators of whether a sexual offender will reoffend or not. Interestingly, the individual measures

which appear to be good at distinguishing recidivists from non-recidivists cover a range of dynamic risk factors for sexual offending, but mainly sit in areas of socio-affective functioning and self-management. The fact that the child offence specific measures did not discriminate recidivists from non-recidivists was not a function of using a mixed group of adult and child offenders. When the differences between groups were examined, using the child molesters from the sample only, there were no differences in scores on these measures between recidivists and non-recidivists.

There was an additional psychometric measure, the impression management subscale of the Paulhus, on which recidivists' scores significantly differed from non-recidivists' scores at both pre- and post-treatment. However, on this measure the non-recidivists scored worse; that is non-recidivists had a significantly higher impression management score both pre- and post-treatment than recidivists. This is the reverse of what was expected. It seems that those who go on to reoffend may not be interested in managing others' impression of them as much as those who do not go on to reoffend. Alternatively they don't have the skills to manage others' impression of them. Either way, these results suggest that impression management may in fact be a protective factor against recidivism. These findings do fit in with previous research which indicates that risk is inversely related to impression management amongst sexual offenders (Mathie & Wakeling, 2011), and that lower impression management scores may be more related to recidivism than higher scores (Mills & Kroner, 2006; Mills, Loza & Kroner, 2003). In line with this, Otter and Egan (2008) also found that self-deception enhancement may protect against antisocial thinking. It would be interesting to examine whether this relationship is affected by levels of denial; future research should attempt to explore this finding further.

In this initial set of analyses there were a few further anomalies that should be mentioned. It is noteworthy, for example, that although not significant when applying the Bonferonni adjustment, the relapse prevention recognition score was actually a positive predictor; that is those who had higher scores on this subscale were more likely to reoffend than those who had lower scores. This suggests that recognising situations where one might relapse may not be important in *preventing* offenders from reoffending. In fact the present findings indicate that recognising these situations may be more harmful than good. Similarly, those who had lower scores on the Children and Sex scale, and those who had higher scores on the Openness to Women scale were more likely to reoffend too. These findings deserve further examination in future research, but they may suggest either that these areas are not as straightforwardly related to recidivism as previously thought, or that the measures used in the present study are not adequately capturing these risk factors.

When the psychometrics were grouped into dynamic risk factor domains and a score for each domain was calculated for each individual, recidivists and non-recidivists' scores significantly differed on all dynamic risk factor domains at the pre-treatment stage. That is, recidivists scored significantly worse than non-recidivists on all of the dynamic risk factor domains (using the OB and the NB measures). At the post-treatment stage fewer dynamic domain scores discriminated recidivists from non-recidivists; the socio-affective functioning and self-management domains however continued to discriminate between groups. These findings indicate that the method used to produce dynamic risk domain scores is functional. These dynamic risk domain scores could be extremely useful for clinicians who are assessing offenders' need for treatment and progression through the system in a resource intensive environment.

When entering these dynamic risk domain scores into regression analyses, the socio-affective functioning domain was a significant predictor both pre- and post-treatment, and added predictive power to static risk alone (when using the domain score calculated from the OB). This domain score continued to add predictive power to static risk when using only those offenders for whom there were full RM2000 scores, which strengthens the support for the predictive power of the domain score. The socio-affective functioning domain focuses on sexual offenders' intimacy deficits (Cortoni & Marshall, 2001). Intimacy deficits have repeatedly been shown to predict recidivism (e.g., Hanson & Morton-Bourgon, 2005) and the present study supports these findings. It is important to note that in the present sample, these problems seem to be better predictors of recidivism than any of the other three domains (sexual interests, offence supportive attitudes, and self management). This could be a function of offenders being more honest about their problems in this domain; offenders may be more likely, for example, to admit their problems surrounding relationships than admit problems with attitudes they hold which are supportive of sexual offending behaviour. It could be that other types of measures used to assess offence supportive attitudes or sexual interests, aside from self-report, would be more associated with recidivism. For example, perhaps behavioural measures of these risk factors, such as the penile plethysmograph, would be better predictors of recidivism.

It is worth noting that when examining the incremental predictive validity of the NB dynamic risk factor domain scores, the individual RM2000 categories were not themselves predictive of recidivism outcome. This could be a function of the smaller sample size used to examine the NB psychometrics. It could also be due to the fact that this group had a shorter follow up period (mean of 2.16 years) compared with those who had been administered the OB

psychometrics (mean of 5.08 years). These analyses should be repeated in a few years when the follow up period has grown for this group of offenders.

The number of dysfunctional dynamic risk factor domains present was also associated with recidivism; the greater the number of dysfunctional dynamic risk factor domains, the greater the recidivism rates were. The pre-treatment PDI score (OB and NB) was also predictive of recidivism outcome and added incremental predictive validity to static risk alone (when using OB PDI score). This indicates that using a measure of PDI at the pre-treatment stage may be a useful classification in terms of who may be most likely to recidivate, and therefore to whom to direct limited resources. The post-treatment PDI score however was not predictive of recidivism.

Overall these findings provide some preliminary support for using psychometric scores to determine presence of risk factors and to help identify treatment need areas. This study supports previous findings that psychometrics can be useful in predicting recidivism and can add incremental validity on top of static risk alone (e.g., Craig et al., 2007; Dempster & Hart, 2002; Thornton, 2002).

One of the notable findings from this study was that in all of the various analyses performed, the pre-treatment psychometric scores seemed to be more predictive than the post-treatment psychometric scores. This supports previous findings (e.g., Hanson & Wallace-Capretta, 2000), and is discussed in detail in subsequent chapters. This could be because the pre-treatment scores more genuinely reflect an individual's propensity to reoffend than the post-treatment scores. The latter will be made up of individuals who have changed different amounts to different extents (e.g., some will remain unchanged on a psychometric, some will

have deteriorated, some will have improved, some will have recovered, and others' scores on a psychometric would have already been within a normative range), making any relationship with an outcome less clear. For example apparent pre- to post-treatment change on a psychometric has different meaning for different individuals, which may cloud the relationship of these scores to recidivism. Some may be faking their responses at the post-treatment stage to appear 'treated'; some may make minimal pre- to post-treatment change on risk factors, whilst others may make a significant amount of change. The pre-treatment scores may therefore be a purer propensity to reconviction than post-treatment scores. This highlights the need for further research to determine whether *change* on these psychometric measures (individual and domain scores) is related to recidivism. This is the topic of interest in Chapter Four.

Limitations

There are a number of limitations of the current study which need to be mentioned. Firstly, as previously discussed, proxy RM2000 scores were used as a measure of static risk, as full scores were not available for the whole sample. Where possible examinations of whether using these proxy static risk scores affected the results were attempted, and the indications were that it did not. However, it would have been better to have been able to use full RM2000 scores as the tool has been validated on a number of different samples and the proxy scores have not. Secondly, the low base rates of sexual recidivism in the present sample meant that the decision was taken to use sexual and violent recidivism as the outcome measure. Had it been possible to use sexual recidivism alone, the results may have been different. It could be, for example, that some measures of sexual interests are only predictive of sexual recidivism and not sexual *and/or* violent recidivism. Thirdly, the fact that the present study used a sample that had completed the OB psychometrics and a smaller sample

that had completed the NB psychometrics complicated the research. It was felt to be important to include the additional NB psychometrics even though the sample size on which they had been administered was smaller, because it would provide additional information. Whilst it did do this, some of the findings do suggest a larger sample size might be needed to detect significant differences when using the new battery.

In line with Walters' (2006) findings, this study suggests that psychometrics which measure constructs related to risk factors can be useful outcome predictors. It seems however that some measures are better predictors of recidivism than others, and particularly that pre-treatment psychometric scores may be better predictors than post-treatment scores. Further research is needed to ascertain whether pre- to post-treatment psychometric change is related to recidivism. Additionally, it seems that those measures targeting the socio-affective functioning dynamic risk domain may be particularly predictive of recidivism, and add incremental validity to static risk.

CHAPTER THREE

IDENTIFYING THE BEST PREDICTORS OF RECIDIVISM USING PROGNOSTIC MODELLING

Chapter Outline

The exploratory analyses conducted in chapter two found that some pre and post-treatment psychometrics could distinguish recidivists and non-recidivists, and that psychometrics have some predictive power both individually and when combined into risk domains. Following these analyses, the present chapter aims to conduct more refined analyses regarding the predictive power of pre and post psychometric measure scores. Specifically, prognostic modelling techniques were employed to examine the best predictors of recidivism in the large sample of sexual offenders which was used for the exploratory analyses in chapter two. The set of predictors examined included available static variables as well as pre and post psychometric (or dynamic) variables.

The aims were:

1) to determine whether static variables or dynamic variables (here measured psychometrically) were more predictive of recidivism outcome, and

2) to determine which of the psychometric variables were the best predictors of recidivism outcome.

The findings indicate that static variables are better predictors or sexual and/or violent recidivism in sexual offenders than dynamic psychometric variables. The psychometric

measuring impulsivity was the best of the psychometric predictors, and an amalgamated score also fared well.

The following article was accepted for publication in Psychological Services, Volume 8, Issue 4, pages 307-318, in 2011.

A further note about the selected methodology and psychometrics used for this chapter is shown in Appendix C along with some additional analyses on absolute risk estimations which were not included in the final version of the paper.

Psychological Services
2011, Vol. 8, No. 4, 307–318

1541-1559/11/$12.00 DOI: 10.1037/a0025516

Identifying Predictors of Recidivism in a Large Sample of United Kingdom Sexual Offenders: A Prognostic Model

Helen C. Wakeling
NOMS Operational Services and Interventions Group, Ministry of Justice, London, United Kingdom, and the University of Birmingham

Nick Freemantle
University College London Medical School

Anthony R. Beech
University of Birmingham

Ian A. Elliott
Lucy Faithfull Foundation, Birmingham, United Kingdom, and the University of Birmingham

This study uses prognostic modeling to identify the best static and dynamic predictors of sexual and violent recidivism for a sample of 3,773 sexual offenders who received treatment in a correctional establishment in the United Kingdom between 1996 and 2006. The use of individual static items was found to predict recidivism better than a modified version of a risk score produced from the static risk assessment Risk Matrix 2000/sexual dimension (RM2000/s) (Thornton et al., 2003). The best predictors of recidivism were age at release, number of sexual appearances, and number of criminal appearances. Pre- and post-psychological measures did not remain in the model in the presence of these three static variables. Further exploratory analyses found that pre-treatment scores on measures related to the socioaffective domain were the most predictive of the dynamic risk domains, but did not add to the predictive power of the static variables. An overall score for deviance was calculated and this score did remain in a model with individual static items. The potential explanations for these findings are discussed along with implications for the assessment of risk in this population.

Keywords: prognostic modeling, sex offenders, recidivism, assessment

CHAPTER FOUR

TREATMENT CHANGE AND ITS RELATIONSHIP TO RECIDIVISM

Chapter Outline

The previous chapters have examined the relationship between pre and post-treatment psychometric scores and recidivism, but not whether *change* as measured by psychometrics is related to recidivism outcome. Thus, this chapter examines the relationship between treatment change, as measured by psychometrics, and sexual and/or violent recidivism. Clinically significant change (CSC) methodology was utilised to produce treatment change scores and groups. CSC methodology has been used in previous studies successfully to examine change within group and individual analyses (e.g., Mandeville-Norden, Beech & Hayes, 2008; Nunes et al., 2011). Analyses revealed that individuals who scored within a 'normal' range before and after engaging in a treatment programme to address their sexual offending had significantly lower reconviction rates than those who scored 'worse than normal' after treatment on a selection of the psychometric measures. These individuals might be more likely to have additional needs, perhaps less offence specific, which might be better met by other types of interventions (such as general thinking skills or cognitive skills programmes). Individuals who were deemed to have 'changed' psychometrically were also less likely to be reconvicted than those who did not change. The analyses indicate that there is some potential use of looking at treatment change as measured by psychometric measures.

The following article was accepted for publication in Psychology, Crime and Law, Volume 19, pages 233-252, in 2013. A further note about the selected methodology and psychometrics used for this chapter is shown in Appendix D.

Psychology, Crime & Law, 2013
Vol. 19, No. 3, 233–252, http://dx.doi.org/10.1080/1068316X.2011.626413

RESEARCH ARTICLE

Investigating treatment change and its relationship to recidivism in a sample of 3773 sex offenders in the UK

Helen Wakeling[a]*, Anthony R. Beech[b] and Nick Freemantle[c]

[a]*Interventions and Substance Misuse Group, NOMS, London, UK;* [b]*Department of Psychology, University of Birmingham, Birmingham, UK;* [c]*Medical School, University of Birmingham, Birmingham, UK*

(*Received 16 June 2010; final version received 13 September 2011*)

This study examines the relationship between psychometric changes in treatment and recidivism in a sample of 3773 sex offenders. All had completed treatment in a prison, between 1996 and 2006. Clinically significant changes were calculated for the psychometrics, and for the overarching psychological problems as represented by the four domains in the Structured Assessment of Risk and Need (Thornton, *Sexual Abuse: A Journal of Research and Treatment*, *14*, 139–154, 2002): (1) sexual interests; (2) pro-offending attitudes; (3) socio-affective problems; and (4) self-regulation problems. Analyses indicate that those whose scores were in the 'normal range' before and after treatment were reconvicted at a significantly lower rate than those whose scores were not in the 'normal range' after treatment on selected psychometric scales. Additionally, participants who were deemed 'changed' overall on three of the four risk domains were reconvicted at a lower rate than those who were deemed not to have changed on these domains. An overall treatment change status was also computed, but this did not add significantly to the predictive validity of a modified version of an actuarial risk assessment tool (RM2000, Thornton et al., *Annals of the New York Academy of Sciences*, *989*, 223–235, 2003) in a Cox regression. The results of the study indicate the potential role and limitations of clinical methodologies in ascertaining whether treatment has worked.

Keywords: sex offenders; risk; treatment; clinical change; recidivism

*Corresponding author. Email: helen.wakeling@noms.gsi.gov.uk

CHAPTER FIVE

SUMMARISING THE FINDINGS SO FAR

This chapter provides a summary of the findings generated so far within this thesis. It was felt useful to produce a summary at this stage to bring together all of the findings from the studies conducted in previous chapters, and at the same time to generate suggestions for further research. In addition, since the previous empirical chapters were conducted, additional findings have been generated in the literature, which it was felt would be useful to summarise. The suggestions for further research at the end of this chapter directly influence the final chapters within the thesis.

The following article was accepted for publication in Aggression and Violent Behavior, Volume 19, pages 138-145, in 2014.

Aggression and Violent Behavior 19 (2014) 138–145

Contents lists available at ScienceDirect

Aggression and Violent Behavior

The relationship between psychometric test scores and reconviction in sexual offenders undertaking treatment

H.C. Wakeling *, G.D. Barnett

Operational Services and Interventions Group, National Offender Management Service, 4th Floor Clive House, 70 Petty France, London SW1H 9EX, United Kingdom

ARTICLE INFO

Article history:
Received 26 June 2013
Received in revised form 7 January 2014
Accepted 9 January 2014
Available online 30 January 2014

Keywords:
Psychometric testing
Risk assessment
Sexual offending

ABSTRACT

The evidence is mixed regarding the utility of psychometric test scores in prediction of the likelihood of sexual reoffending. This paper summarizes the research examining the relationship between psychometric measures and sexual recidivism, before detailing the findings of four large-scale studies in England and Wales, and comparing the findings of these studies to similar studies from other countries. The implications of the evidence to date are discussed, and recommendations are made for the future of psychometric testing as a way of determining risk of recidivism, dynamic risk factors, change over treatment, and the efficacy of rehabilitative programs.

Contents

CHAPTER SIX

DEVELOPMENT OF A NEW SET OF MEASURES OF RISK OF SEXUAL OFFENDING

Chapter Outline

The study presented in this chapter used pre and post-treatment psychometric assessment scores for a sample of 2,836 offenders who undertook sexual offender treatment in custody in England and Wales. Factor analysis was used to generate new domain measures of: offence-supportive attitudes, socio-affective functioning and self-management, from the psychometrics currently used within Sex Offender Treatment in custodial settings. The reliability of these new measures was found to be good. The battery was then examined using a subsample for whom reconviction data was available; pre- and post-socio-affective functioning, and pre- and post-self-management domain scores significantly predicted general, and sexual and/or violent recidivism. An overall pre- and post-treatment psychometric score was also calculated. The post-treatment score added to the prediction of general recidivism, but these overall scores were not able to predict sexual or sexual and/or violent recidivism. The findings suggest that these new measures are reasonably successful in the identification of sexual offender recidivists, and hence are indicative of management and treatment need. The limitations of this research are discussed along with suggestions for future research.

Introduction

Thus far the NOMS SOTP pre and post-treatment psychometrics have been examined in terms of their relationship with recidivism. The results of the studies conducted so far were

summarised in the previous chapter. The main findings have been that psychometric test scores are moderately predictive of future reconviction but not more predictive than static factors. In addition, pre-treatment psychometrics have generally been found to be more predictive than post-treatment psychometrics, and there is some indication that psychometrics clustered into risk domain scores may enhance their predictive power. These findings suggest therefore that psychometric measurement of dynamic risk domains, rather than of individual dynamic risk factors, could potentially add meaningfully to the assessment of risk of sexual offenders. However, the lack of an association between post-treatment scores and subsequent reconviction, and the lack of incremental validity to static risk assessment, poses a problem for those involved in the provision of treatment for offenders, and in particular the use of psychometrics in helping to identify dynamic risk factors within a treatment context. In the current fiscal climate where resources are stretched it seems advantageous to only be using those psychometrics which have a good predictive power, and/or using fewer or amalgamated scores if appropriate.

As a result of the findings reported in chapters 2 through to 5, the current battery of psychometric tests (NB) used as part of the National Offender Management Service's (NOMS) sexual offending treatment in the UK, was examined in the present study. This study aimed to create a more refined battery that measures risk domains, and to test the predictive validity of these new risk domain measures. The importance of both reliability estimates and face validity of the new assessment was fundamental, and the intention of the study was to try to examine the item pool and unpack the multiple components underpinning some of the scales, leading to a shorter and more focussed assessment. As well as addressing concerns about artificially inflated reliability statistics (Cortina, 1993), this would also reduce the level of time and therefore resources required to administer the tests. The ultimate aim therefore

was to reduce the item pool of the current psychometric battery, and to produce new risk domain measures, that are reliable and valid, whilst retaining or enhancing the efficiency and predictive power of the measures. Therefore, the present research aimed to answer the following research questions:

1) Can a shorter battery be developed that assesses risk domains?
2) Is such a system reliable and does it have good validity?

Method

Participants

A sample of 2836 adult male sexual offenders, who had taken part in a sex offender treatment program in one of 28 prisons across England and Wales between 2003 and 2010, was used in the study. All participants had committed at least one offence with a sexual element. Treatment dropouts were not included in the sample due to the fact that only data for those who completed both pre and post-treatment psychometric data were collected centrally in the UK. This is a smaller sample from the samples used in the previous three empirical chapters, as it uses only those sexual offenders who have been administered the NB psychometrics since 2003. The mean age of the sample was 46.37 (SD 12.61, range 20-84). The majority of the sample had a sentence length of over four years. Eighty-one percent of the sample were White, the rest were minority ethnic. The majority of the sample had an index offence of a sexual nature (87%), most of the remaining 13% had an index offence of a violent nature, but either had a previous conviction for a sexual offence, or their index violent offence had a sexual component. The sample was divided into offence categories (child molester/rapist) based on the age of the victim of their most recent offence. While it is likely that some offenders in the sample have offended against both adults and children in their offending history, this was the best available way of determining offender type for the purpose of this

study. Of the total sample, 56% were child molesters, and 44% were classified as rapists (defined as those with an adult female victim of 16 years or over).

Reconviction data was available for 38% of the sample (N=1080), and was sourced from the Ministry of Justice Police National Computer (MoJPNC) in the UK. This data was only available for a proportion of the sample due to the fact that the remainder of the sample had either not been released from prison yet or had not been released for long enough to obtain reconviction data (6 month buffer period used). It was felt best to develop the new measures using the whole sample as it is generally accepted that you should try to get the largest possible sample when conducting a factor analysis (Henson & Roberts, 2006), but only the smaller subsample of those with reconviction data were used when examining the predictive power of the newly generated measures. General, sexual, and sexual and/or violent recidivism of the subsample was examined. The average length of follow up was 799 days (2.19 years) for those whose reconviction data were available; the two-year sexual reconviction rate was 1.5%; the two year sexual and/or violent reconviction rate was 3.3%; and the two-year general reconviction rate was 9.6%. The sample with reconviction data was used to examine the predictive ability of the newly devised assessment and associated domain scores.

Psychometric measures

A number of measures used to assess three of the four risk domains pre- and post-treatment on the current Sex Offender Treatment Program (SOTP) used in UK prisons were utilized for this study (those included in the NB, as described in chapter two). The psychometric tool used to measure the Sexual Interests domain was not examined in this study because a scale measuring the factors in this domain, the My Private Interests Scale, has already been

developed and examined elsewhere (Farren & Barnett, 2014; Williams, 2007). This measure has already replaced the longer Multiphasic Sex Inventory (Nichols & Molinder, 1984), which is currently utilized to assess risk factors in this domain in the UK, but the data for this measure was only available for a very small proportion of the current sample, and thus was not included in the present study. A number of other psychometric measures (used currently as part of the NOMS SOTP psychometric battery: NB) were not used within the present study according to the following criteria:

1) Any psychometric measure or subscale which was deemed not to directly measure a dynamic risk factor within SARN was omitted. The following were omitted for this reason: Minimisations, Paulhus Self-Deception Scales, the Fantasy and Personal Distress subscales of the Interpersonal Reactivity Index, and the Rehearsal and Emotional inhibition subscales of the Emotion Control Questionnaire.

2) Any psychometric measure which was copyright was not used in the present research as the intention was to use the results of this study to produce a new and refined psychometric battery for SOTP. Items from copyright measures cannot be used in this way. The following were omitted for this reason: The Social-Problem Solving Inventory-Revised.

3) Any psychometric measure found to be unrelated to recidivism from the previous chapters was omitted. The following were omitted for this reason: The Relationships Style Questionnaire subscales, the Hypermasculinity Inventory subscales.

4) Any psychometric measure which was removed from the psychometric battery in later revisions (post 2003) was omitted. The following were omitted for this reason: The Relapse Prevention questionnaire.

Hence, those psychometrics that are not copyright, that have shown some relationship with recidivism, that are in the current SOTP psychometric battery, and that target risk factors

within the remaining three risk domains, offence-supportive attitudes, socio-affective functioning, and self-management were used. The measures that were used are listed below (and are described in detail in Chapter Two). Decisions about which scales and subscales measure each of the domains were made by examining the face validity of the measures.

Offence-supportive attitudes domain measures

The Entitlement to Sex scale (Hanson, Gizarrelli & Scott, 1994)

The Women are Deceitful Scale (NOMS, unpublished)

The Children and Sex Questionnaire (Beckett, 1987): The cognitive distortions subscale items were used for the offence-supportive attitudes domain (whereas the emotional congruence subscale was used to measure socio-affective functioning).

The Sex with Children is Justifiable Questionnaire (Marshall, 1995; Mann, Webster, Wakeling & Marshall, 2007)

Socio-affective functioning domain measures

The Revised Dissipation Rumination scale (Caprara, 1986; Wakeling & Barnett, 2011)

The University of California Loneliness Scale (UCLA; Russell, Peplan & Cutrona, 1980)

The Self-Esteem scale (Thornton, Beech & Marshall, 2004; Webster, Mann, Thornton & Wakeling, 2007)

The Interpersonal Reactivity Index (IRI; Davis, 1980): The Emotion Concern and Perspective Taking subscales of this measure were used.

Locus of Control (Levenson, 1974)

The Openness to Men and Women Scales (Underhill, Wakeling, Mann & Webster, 2008)

The Children and Sex Questionnaire (Beckett, 1987): The Emotional Congruence subscale of the Children and Sex questionnaire, described earlier.

Self-management domain measures

The Impulsivity Scale (Eysenck & Eysenck, 1978)

The Emotion Control Questionnaire (Roger & Najarian, 1989): The Aggression Control and Benign Control subscales were used.

Risk of Sexual Reconviction

Risk of sexual reoffending was calculated for each individual using Thornton's Risk Matrix 2000/ sexual scale (RM2000/s; Thornton et al., 2003), as described in previous chapters.

Risk of General Reconviction

The revised Offender Group Reconviction Scale (OGRS3, Howard, Soothill, Francis & Humphreys, 2009) was employed to measure risk of general reconviction. This is an actuarial predictor of general reoffending, used extensively in assessments completed by probation and prison staff in the UK. It estimates the percentage likelihoods of proven reoffending (any conviction or caution for a new offence) committed within one and two years of the start of a community sentence or discharge from custody. It includes only static risk factors: age, sex and criminal history. The criminal history factors include a 20-group classification of current offence type, an offending 'rate' based on the number of sanctions (convictions and cautions) in the offender's criminal history and the number of years between first and current sanction, and an additional variable identifying those with very short criminal histories.

Data Analytic Plan

1. Generating New Scales

Pre- and post-treatment psychometric test scores on the scales reported above were routinely collected centrally by the Operational Services and Interventions Group (OSIG), National Offender Management System, UK. For the purpose of production of the new psychometric

tests only pretreatment psychometric data were used following conventions of previous research (e.g., Wakeling, 2007). When testing the scales with the smaller subsample those who had reconviction data both pre- and post-treatment data were used. Pretreatment psychometric data were first split into the three risk domains (through examination of their face validity), and then a series of factor analyses were conducted by risk domain.

Factor analysis is one of the most widely used techniques in the development and evaluation of psychological measures used to examine validity (Floyd & Widderman, 1995). Exploratory factor analytical approaches are used to identify underlying dimensions of a set of variables, or to reduce data to summary indices (Bartholomew, Steele, Moustaki & Galbraith, 2002). Confirmatory approaches on the other hand, are used to confirm a priori hypotheses or structures (Bentler, 1989; Jöreskog & Sörbom, 1989). Exploratory factor analyses techniques were used for the present study.

For all analyses, items were recoded so that a higher score indicated a criminogenic problem.

Offence Supportive Attitudes (OSA) Domain. First, all of the items relating to attitudes towards children, within those measures listed above, were subject to factor analysis. These items were examined using a child abuser population only. The remaining scales (Entitlement to Sex, and Women are Deceitful), were then subject to a second factor analysis, using only those in the sample whose offences were against adults (defined as victims of 16 years of age or over). The decision was made to conduct two factor analyses on the offence supportive attitudes items due to the fact that there was a logical split between those items looking at the attitudes supporting offending against women and those supporting offending against children.

Socio-affective Functioning Domain. Each item in the measures listed in the Socio-affective Functioning domain was entered into one of two separate factor analyses. The first contained those scales, which currently use a dichotomous response format (Self-esteem and Ruminations scales). A Jaccard correlation matrix was used for this analysis due to the dichotomous nature of the scales. The second contained the remaining measures (UCLA Loneliness scale, Empathic Concern and Perspective Taking subscales of the IRI, Locus of Control scale, Openness to Men and Women scale and the Emotional Congruence with Children subscale of the Children and Sex Questionnaire). The UCLA was rescaled to share the same response format as the other measures used in this analysis.

Self-Management Domain. The impulsivity items, and the ECQ Benign Control and Aggression Control subscales were entered into one factor analysis.

For all sets of analyses an Exploratory common Factor Analysis (EFA) was performed on the data using Maximum Likelihood Extraction (MLE) techniques, with oblique rotations. Velicer's minimum average partial (MAP) test, a validated procedure for determining the number of components in factor analyses (O'Connor, 2000), was utilised as the main method to select the number of factors to extract in the analyses. In addition, other common methods of selecting the number of factors were used including examination of the scree plot (Cattell, 1966), and the Eigenvalue greater than one rule (Kaiser, 1960). Zwick and Velicer (1986) in an examination of the accuracy of various methods of factor retention found the minimum average partial method to be one of the most accurate methods, but the current accepted guidance is to use multiple criteria and reasoned reflection (Henson & Roberts, 2006). In addition, prior to analyses, it was decided to only retain items with a loading of greater than

.5, in accordance with Costello and Osborne (2005), unless there was a strong reason to do otherwise.

2. Examining the Predictive Validity of the Newly Devised Scales

Total domain scores were computed for the three SARN domains examined in the present study. Once these new scales were created, their discrimant and predictive validity were examined. The data analyses for testing the validity of these scales proceeded in a number of stages.

1) First, pre and post scores on the new measures are presented along with a measure of effect. These effect size magnitudes are presented using Cohen's *d*; as such the pre to post treatment change is presented in standard deviation units, with values of .20, .50 and .80 corresponding to small, medium and large effect sizes respectively (Cohen, 1992). This provides information on a) differences on scale scores between recidivism groups, and b) the level of treatment change observed on the newly devised measures.

2) Second, t-tests were computed to compare scores on the new scales between those who had been reconvicted and those who had not to examine discriminant validity. This provides some indication of whether the scales are sensitive to differences between groups whose scores would be expected to differ (e.g., recidivists should score worse on all measures pre and post-treatment in comparison with non-recidivists).

3) Third, the predictive accuracy of the new measures (both pre and post-treatment) was examined using receiver operating characteristic (ROC) analyses and Cohen's *d*. The ROC analyses generate an area under the curve (AUC) value ranging from 0 to 1.0, which provides

a likelihood statistic that a randomly selected recidivist scores worse on the measures than a randomly selected non-recidivist. AUCs of .5 represent likelihood no better than chance, whereas an AUC of 1.0 represents perfect accuracy. The interpretation suggested by Douglas et al. (2008), as described in chapter two, was again utilised here (values of .71 representing a large effect size and .64 a moderate effect size). Although some argue that there is an overreliance on AUCs as a measure of effect size, and that it is misleading to use conclusions based on AUCs to make decisions about individuals (Cooke & Mitchie, 2014), they are currently the most commonly used statistic in the field of risk assessment and rehabilitation (e.g., Hanson, Helmus & Thornton, 2010). Cohen's *d* was also used to represent an effect for the pre and post treatment measures, with values of .20, .50 and .80 corresponding to small, medium and large effect sizes, respectively (Cohen, 1992).

4) Fourth, the predictive accuracy of the pre-post treatment raw change score and the standardized residual change scores (RCZs) were then examined using receiver operating characteristic (ROC) analyses and Cohen's *d*. The AUC and Cohen's *d* values as described above were used again for these analyses. Following on from the work of Beggs and Grace (2011) and Olver et al. (in press), RCZs were computed by regressing the raw psychometric change scores on the pre-treatment score for each measure. This residual is then standardised and the association between this and the outcome is examined. The use of RCZs attempts to examine the effect of the pre to post-treatment change whilst controlling for an individual's starting point (the pre-treatment score).

5) Next a series of nested Cox regression survival analyses were conducted to examine the ability of the new measures in predicting reconviction outcomes. These nested models used the pre-treatment score as the baseline model following on from both findings in earlier

chapters of this thesis that pre-treatment measures were better predictors of outcome than post-treatment measures, and from other similar research findings (e.g., Hanson & Wallace-Capretta, 2000; Quinsey et al., 1998). It was important to make this decision based on a range of evidence, rather than just the evidence produced from this thesis, to overcome tautological issues, as similar samples are utilised for the various studies within this thesis. The predictive ability of pre-treatment, post-treatment and change scores were examined within these nested Cox regression analyses, and AIC values were compared to determine the best fitting models for each measure and each reconviction outcome. This methodology therefore allows for the examination of whether the change score has a unique contribution to the prediction of reconviction not accounted for by the pre-treatment score. This is a similar methodology to examination of the standardised residual change scores (RCZs) as conducted by Beggs and Grace (2011) and described in detail in the previous chapter, and allows for an examination of the predictive power of the change score whilst taking into account the pre-treatment score.

6) Lastly, an overall deviancy score was calculated by summing the standardised scores from all revised scales. Further Cox regressions were conducted to examine the predictive power of this score alongside static risk assessment tools.

Results

1. Generating New Scales

Offence-supportive attitudes domain: Items relating to attitudes towards children

A factor analysis using varimax rotation of the two scales relating to attitudes towards children was conducted (item pool, *N* =33). For this analysis, only the child abusers in the current sample were included. The analyses suggested a three factor solution (based on Velicer's minimum average partial (MAP) test). The scree plot supported extraction of three factors whilst the Eigenvalue rule suggested four factors should be extracted. On examination of the three and four factor solutions, the three factor solution was selected as the most meaningful. Two items with loadings of less than .5 were not retained in these three subscales. The table below shows the items and factor loadings. Factor 1 contained 17 items relating to beliefs that sex with children was not harmful to them and hence was entitled 'Denial of Harm'. Factor 2 contained nine items relating to beliefs about children being sexual beings, wanting to have sex with adults, and leading adults on sexually. This factor was thus given the title 'Child as Sexual'. The final factor contained five items relating to beliefs that normalise the sexual contact between children and adults, entitled 'Normalising Sexual Contact with Children'. The Cronbach's alpha was examined to test the internal consistency of these scales. An alpha of .95 was generated for the first factor (*N* =1166, *n*=17), .88 for the second factor (*N*=1177, *n*=9), and .79 for the third factor (*N*=1187, *n*=5). The overall scale, putting all three of these generated subscales together produced an alpha of .94 (*N*=1137, *n*=31).

Table 6.1: Factor Loadings for the Items Relating to Sexual Interest Domain in Children (*N*=31)

Item	Factor 1 Loading	Factor 2 Loading	Factor 3 Loading
SWCH16 Many children benefit from having sex with an adult	**.785**		
SWCH15 Having sex with a child is a good way to teach them about sexuality	**.776**		
SWCH17 It is ok to have sex with a child as long as you don't force the child into it	**.752**		
SWCH9 It is far better for young people to have their first sexual experience during childhood with an adult than to risk what is sure to be an unpleasant experience with someone their own age when they are a teenager	**.731**		
SWCH11 Having sex with a child is not really all that bad because it doesn't really harm the child	**.708**		
SWCH8 Sex between adults and children is quite natural and healthy and it is only because of the repressive rules of society that men are punished for doing this	**.693**		
SWCH18 Nowadays it is not so bad to have a sexual relationship with someone who is under-age because kids know so much more about sex than they used to	**.688**		
SWCH14 Having sex with a child is a way of expressing love and affection for that child	**.685**		
SWCH2 Most children actually enjoy sex with an adult so long as the man doesn't hurt them	**.674**		
SWCH10 A man can't help having sex with a child if the child acts in a provocative manner	**.666**		
SWCH13 Children are old enough to decide whether or not they want to have sex with someone	**.648**		
SWCH4 Children enjoy sexual attention from adults	**.629**		
SWCH7 Children who are unloved by their parents are actually helped by men who have sex with them	**.616**		
SWCH6 Children who do not wear underwear and who sit in a way that is revealing are suggesting sex	**.609**		
SWCH1 Children actually enjoy sex with a man if the man is nice to them	**.603**		
SWCH12 If an adult has sex with a child who enjoys it and seems to want it, it shouldn't be considered a crime	**.601**		
SWCH5 Men who have sex with children are usually led into it by the child	**.597**		
CAS10 Children can lead adults on		**.774**	
CAS9 Children can flirt with adults		**.709**	
CAS12 Children can lead adults astray		**.706**	
CAS14 People underestimate how much children know about sex		**.688**	
CAS5 Children are not as innocent as most people think		**.678**	
CAS15 Some children could teach adults about sex		**.623**	
CAS11 Children sometimes ask adults for sex		**.615**	
CAS1 Children know a lot about sex		**.586**	
SWCH3 Many children are sexually seductive towards adults		**.548**	
CAS13 There is no harm in sexual contact between children and adults			**.688**
CAS4 There is nothing wrong with sexual contact between children and adults			**.664**
CAS7 If children want, they should be allowed to have sexual relationships with adults			**.624**
CAS8 Most sexual contact between adults and children does not cause any harm			**.529**
CAS3 Children want sexual contact with adults			**.513**
Items not included			
CAS6 When adults and children have sexual relationships it is not always the adults fault	.124	.413	.494
CAS2 Children know more about sex than adults	.272	.389	.465

Note: SWCH = Sex with Children questionnaire; CAS = Children and Sex Questionnaire

Offence-supportive attitudes domain: items relating to attitudes towards women. A factor analysis using varimax rotation of the two scales relating to attitudes towards women was conducted (item pool, *N* =14). For this analysis, only the offenders with adult victims were included. The Velicer MAP test and other methods all suggested a one factor solution. Five items with loadings of less than .5 were not retained. Table 6.2 shows the items and factor loadings. The Cronbach's alpha was examined to test the internal consistency of this scale; an alpha of .83 was generated (*N*=1489, *n* =9).

Table 6.2: Factor Loadings for the Items Relating to Negative Attitudes towards Women (*N*=9)

Item	Factor 1 Loading
WAD5 Many times a woman appears to care but just wants to use you	**.790**
WAD3 I feel that many times women told the truth but now I know otherwise	**.764**
WAD1 When it comes down to it many women are deceitful	**.704**
WAD4 I used to think that most women told the truth but now I know otherwise	**.690**
ENT8 A man who is denied sex suffers more than a women who has sex when she does not want to	**.628**
ENT3 Women should oblige men's sexual needs	**.593**
ENT9 A person should have sex whenever it is needed	**.572**
ENT4 Men need more sex than women do	**.547**
WAD2 The women in my life have let me down	**.511**
Items not included	
ENT5 I have a higher sex drive than most people	.480
ENT6 I am often bothered by thoughts about having sex	.439
ENT1 Everyone is entitled to sex	.405
ENT7 I have no trouble going without sex if my partner is not interested *	.377
ENT2 Sex must be enjoyed by both parties *	.193

Note: WAD = Women as Deceitful scale, ENT = Entitlement to Sex scale.
** Items are reverse scored*

Socio-affective functioning domain: all items relating to risk factors within this domain. Due to the different response formats of the measures, two separate factor analyses of all of the items within this risk domain were conducted (total item pool, *N*=107). The first examined those items from the measures using dichotomous response formats; the Short Self-Esteem and Ruminations scales (item pool, *n*=23). The second examined those items from the measures that use a Likert scale: Locus of Control scale, Empathic Concern and

Perspective Taking subscales of the IRI, UCLA Loneliness scale, the Openness to Women and Men scale, and the Emotional Congruence with Children subscale of the Children and Sex Questionnaire (item pool, *n*=84).

A factor analysis using varimax rotation (using the Jaccard matrix appropriate for items with dichotomous response scales) on the Self-esteem and Ruminations measures extracted two main factors which corresponded directly with a Self esteem measure and a Ruminations measure. Loadings of less than .5 were not included in this factor (five items were removed). The factor loadings are shown in Table 6.3. The Cronbach's alpha was examined to test the internal consistency of these two subscales. An alpha of .85 was generated (*N*=2704, *n* =7) for the new self esteem scale, and an alpha of .69 (N=2743, *n*=5) was generated for the new ruminations scale. The ruminations subscale had an unacceptable Cronbach's Alpha according to criteria set out by Kline (2000), who suggest that a good scale must have an Alpha of at least .70. Therefore the Cronbach Alpha of the subscale was calculated with additional items added which had factor loadings of just below .50 (added in one by one according to their factor loadings). The internal consistency of the scale was improved to .73 (*N*=2699, *n*=10) when 10 items were included in this subscale. The lowest factor loading for these items was .40. The 10 item subscale was chosen as preferable to the originally generated five item subscale.

Table 6.3: Factor Loadings for the Dichotomous Items within the Socio-affective Functioning Domain (*N*=14)

Item	Factor 1 Loading	Factor 2 Loading
SE7 I am pretty happy with the way I am *	**.795**	
SE8 I have a low opinion of myself	**.792**	
SE2 I like the sort of person I am *	**.789**	
SE4 I understand myself *	**.685**	
SE6 Things are all mixed up in my life	**.660**	
SE3 I often feel ashamed of myself	**.605**	
SE1 I often wish I were someone else	**.590**	
RUM7 I hold a grudge for a very long time against people who have offended me		**.728**
RUM5 I always retaliate eventually if somebody has wronged me		**.588**
RUM3 I forgive easily when I have been wronged *		**.586**
RUM2 I get more satisfaction from revenge if I have had to wait for it		**.569**
RUM6 I feel so strongly about some of the wrongs that have been done to me that I won't accept any excuses		**.545**
RUM11 If someone harms me, I am unable to relax until I have retaliated		**.493**
RUM4 It takes me many years to get rid of a grudge		**.457**
RUM15 I sometimes help those who have done me wrong *		**.430**
RUM12 when something outrages me, thinking about it only makes me angrier		**.424**
RUM14 I often find it difficult to get to sleep because I can't stop thinking about a wrong done to me		**.404**
Items not included		
SE5 I think I can make a success of my life *	.443	.333
RUM13 I am often sulky	.424	.333
RUM9 I can remember very well the last time I were insulted	.134	.389
RUM10 I still remember the wrongs that I have suffered even after many years	.199	.373
RUM1 I will always remember the injustices I have suffered	.236	.366
RUM8 I remain aloof from people who annoy me in spite of any excuses	.123	.321

Note: SE = Short Self-Esteem scale, RUM = Ruminations scale
** Items are reverse scored*

The second factor analysis was conducted on the remaining measures within this domain from a total pool of 84 items. The Velicer MAP test suggested a twelve-factor solution, whilst the Eigenvalue method suggested a 16 factor solution, and the scree test method suggested a five or six factor solution. To determine which of these factor solutions should be selected, each of the different solutions were scrutinised in terms of meaning using reasoned reflection. Those solutions with a high number of factors (12 and 16) were meaningful to an extent when looking at the items, but many of the emergent factors had a very small number of items (indicative of weak factors). It was decided therefore to select either a five or a six factor solution (based on the scree test). The five factor solution was selected as the more meaningful of the two solutions scrutinised. The five factors in this solution almost exactly

reflected the measures, which were entered and can be described as follows: loneliness, openness to men and women, emotional congruence with children, locus of control, and empathic concern/perspective taking. Using only those items with a loading > .5; in total 17 items were omitted. The factor loadings are shown in Table 6.4. The Cronbach's alpha of these factors were examined to test their internal consistency. The following alphas were generated for each of the scales: loneliness .95 (*N*=2171, *n* =18); openness to men and women .89 (*N*= 2663, *n* =16); emotional congruence with children .88 (*N*= 2579, *n* =15); locus of control .82 (*N*= 2768, *n* =9); and empathic concern and perspective taking .83 (*N*=2782, *n* =9).

Table 6.4: Factor Loadings for the Items within the Socio-affective Functioning Domain with Likert Response Formats

Item	Factor 1 Loading	Factor 2 Loading	Factor 3 Loading	Factor 4 Loading	Factor 5 Loading
Loneliness 15 I felt isolated from others	**.837**				
Loneliness 4 I felt alone	**.817**				
Loneliness 11 I felt left out	**.789**				
Loneliness 7 I was no longer close to anyone	**.775**				
Loneliness 3 I could not find people to turn to	**.772**				
Loneliness 13 Nobody knew me well	**.753**				
Loneliness 17 There were people who really understood me *	**.751**				
Loneliness 19 There were people I could turn to *	**.744**				
Loneliness 14 There were people I could talk to *	**.738**				
Loneliness 8 No one around me shared my interests and ideas	**.737**				
Loneliness 5 I felt part of a group of friends *	**.716**				
Loneliness 2 I lacked companionship	**.715**				
Loneliness 6 I had a lot in common with the people around me *	**.704**				
Loneliness 12 My social relationships were superficial	**.685**				
Loneliness 16 I could find companionship when I wanted it *	**.675**				
Loneliness 18 People were around me but not with me	**.674**				
Loneliness 1 I felt in turn with the people around me *	**.623**				
Loneliness 9 I was an outgoing person *	**.550**				
OPW3 I find it easy to make friends with women *		**.663**			
OPW5 I have always found it difficult to talk to women		**.638**			

OPW4 Women find me easy to make friends with *	**.635**			
OPM9 Men find me easy to make friends with *	**.631**			
OPM4 I can discuss my problems with men *	**.618**			
OPM5 I have always found it difficult to talk to men	**.610**			
OPW6 I have always been very shy with women	**.609**			
OPM6 I have always been very shy with men	**.609**			
OPW9 I find it hard to talk to women about the things that really matter to me	**.605**			
OPW2 Women generally like me *	**.586**			
OPW7 I can discuss my problems with women *	**.567**			
OPM7 Men generally like me *	**.567**			
OPM2 I like spending my time chatting to men *	**.538**			
OPM3 I find it hard to talk to men about things that really matter to me	**.529**			
OPM1 I find it easy to make friends with men *	**.529**			
OPW8 I never let women see what I'm really like	**.508**			
CAS18 Thinking about children makes me feel good		**.691**		
CAS28 When I feel low children cheer me up		**.681**		
CAS21 Children stop me feeling lonely		**.678**		
CAS16 I prefer to spend my time with children		**.634**		
CAS24 I feel more comfortable with children than with adults		**.615**		
CAS22 Children are special for me		**.613**		
CAS30 Children seem to seek me out		**.604**		
CAS27 I am better than most people at getting along with children		**.588**		
CAS25 Sometimes I meet a child who I know has special feelings about me		**.576**		
CAS19 I know when children are interested in me		**.567**		
CAS26 I am better than most people at understanding children		**.566**		
CAS20 Sometimes children look at me in a special way		**.554**		
CAS29 Some children prefer to be with me rather than their parents		**.554**		
CAS17 I have loved a child at first sight		**.550**		
CAS23 Children remind me of myself		**.527**		
LOC3 Everyone knows that luck or chance determines the future			**.600**	
LOC12 When I am under stress, the tightness in my muscles is due to things outside my control			**.595**	
LOC10 I believe people are victims of circumstances beyond their control			**.592**	
LOC9 My life is controlled by outside actions and events			**.576**	
LOC2 A great deal of what happens to me is just a matter of chance			**.571**	
LOC17 In my case maintaining control over my problems is mainly due to luck			**.564**	
LOC18 I have often been blamed for events beyond my control			**.556**	
LOC6 My problems will dominate my life			**.520**	
LOC14 It is impossible to control irregular fast breathing when I am having difficulties			**.520**	
PT3 I sometimes try to understand my friends better by imagining how things look from their perspective *				**.661**

PT5 I believe that there are two sides to every question and I try to look at them both *				**.633**
PT2 I try to look at everybody's side of a disagreement before making a decision				**.630**
PT7 Before criticising somebody, I try to imagine how I would feel if I were in their place *				**.623**
EC3 When I see someone being taken advantage of, I feel kind of protective towards them *				**.614**
EC6 I am often quite touched by things that I see happen *				**.597**
PT6 When I'm upset with someone, I usually try to 'put myself in his shoes' for a while *				**.595**
EC1 I often have tender, concerned feelings for people less fortunate than me *				**.592**
EC7 I would describe myself as a pretty soft-hearted person *				**.536**
Items not included				
Loneliness 10 There were people I felt close to *	.451			
OPM8 I never let men see what I'm really like		.478		
OPW1 I like spending my time chatting to women *		.436		
LOC4 I can control my problems only if I have outside support			.488	
LOC11 To continually manage my problems I need professional help			.468	
PT4 If I'm sure I'm right about something, I don't waste much time listening to other people's arguments			.429	
PT1 I sometimes find it difficult to see things from other guy's point of view			.363	
LOC8 Becoming a success is a matter of hard work, luck has little or nothing to do with it *			.233	
LOC13 I believe a person can be master of his own fate *			.228	
EC5 When I see someone being treated unfairly, I sometimes don't feel much pity for them				.423
LOC1 I can anticipate difficulties and take action to avoid them *				.403
EC4 Other people's misfortunes do not usually disturb me a great deal				.350
LOC16 I am confident of being able to deal successfully with future problems *				.349
EC2 Sometimes I don't feel very sorry for other people when they have problems				.306
LOC5 When I make plans I am almost certain I can make them work *				.257
LOC15 I understand why my problems vary so much from one occasion to another *				.245
LOC7 My mistakes and problems are my responsibility to deal with *				.209

Note: OPW = Openness to Women scale, OPM = Openness to Men scale, CAS = Children and Sex Questionnaire, Emotional Congruence with Children subscale, LOC = Locus of Control scale, PT = Perspective Taking and EC = Empathic Concern subscales of the Interpersonal Reactivity Index.
* Items are reverse scored

A total socio-affective functioning score was also computed based on scores derived from the seven new scales in this domain (including the self esteem and ruminations scales). The internal consistency of this total socio-affective functioning scale was .91 (*N*=1753, *n* =84).

Self-management domain: All items relating to risk factors within this domain. One factor analysis was conducted on all of the items within this domain (total item pool, *N*=41). A two factor model emerged as the best fitting according to two of the three criteria of factor selection used (the Eigenvalue > 1 criteria suggested a greater number of factors which were deemed less meaningful upon examination), omitting a total of 21 items with loadings of less than .5 (two items which could be rounded up from .499 to .5 were retained). Factor 1 was comprised of 14 impulsivity items, thus entitled 'Impulsivity'. Factor 2 was comprised of Aggression items, thus entitled 'Aggression'. Factor loadings are shown in Table 6.5. The internal consistency of the two subscales were .90 (*N*=2656, *n*=14), and .67 (*N*=2735, *n*=6), respectively. The Aggression subscale had an unacceptable Cronbach Alpha according to Kline's (2000) criteria. Therefore, the Cronbach Alpha of the subscale was calculated with additional items added which had factor loadings of just below .50 (added in one by one according to their factor loadings). The internal consistency of the scale was improved to .70 (*N*=2703, *n*=9) when nine items were included in this subscale. The lowest factor loading for these items was .47. The nine item subscale was chosen as preferable to the original six item subscale. The internal consistency of the total self management scale (including both the Impulsivity and Aggression subscales) was .87 (*N*=2590, *n*=23).

Table 6.5: Factor Loadings relating to items from the Self-Management Domain (*N*=23)

Item	Factor 1 Loading	Factor 2 Loading
Impulsivity1 Do you generally do and say things without stopping to think?	**.761**	
Impulsivity2 Do you often get into a jam because you do things without thinking?	**.750**	
Impulsiveness3 I often do or say things I later regret	**.719**	
Impulsivity13 Do you usually think carefully before doing anything? *	**.709**	
Impulsivity4 Do you often do things on the spur of the moment?	**.693**	
Impulsiveness6 I often say things without thinking whether I might upset others	**.689**	
Impulsivity8 Do you get so 'carried away' by new and exciting ideas that you never think of possible snags?	**.652**	
Impulsivity6 Do you often get involved in things you later wish you could get out of?	**.652**	
Impulsivity11 Before making up your mind do you consider all the advantages and disadvantages? *	**.648**	
Impulsivity5 Do you mostly speak before thinking things out?	**.625**	
Impulsiveneness9 Almost everything I do is carefully thought out	**.586**	
Impulsiveness5 I frequently change my mind about things	**.540**	
Impulsiveness14 I sometimes just come out with things that embarrass people I'm with	**.527**	
Impulsiveness4 I'm not easily distracted *	**.523**	
Aggression1 If someone pushed me, I would push back		**.567**
Aggression6 If someone were to hit me, I would hit back		**.557**
Aggression9 If I see someone pushing into a queue ahead of me I usually just ignore it *		**.552**
Aggression3 I've been involved in many fights or arguments		**.507**
Aggression2 No-one gets one over on me – I don't take things lying down		**.499**
Aggression11 I lose my temper quickly		**.498**
Aggression13 If a friend borrows something and returns it dirty or damaged, I usually just keep quiet about it *		***.484***
Aggression14 If someone insults me I try to remain as calm as possible *		***.474***
Aggression4 If I'm badly served in a shop or restaurant I don't usually make a fuss		***.469***
Items not included		
Impulsiveness7 My interests tend to change quickly	.495	.151
Impulsiveness13 I like planning ahead rather than just seeing how things turn out *	.488	.064
Impulsviness2 I often take chances crossing the road	.485	.218
Impulsivity7 Do you generally 'look before you leap'? *	.482	.038
Impulsivity12 Do you usually make your mind up quickly?	.446	.086

Impulsiveness11 I can't stand having to wait for anything	.424	.284
Impulsivity10 Are you often surprised at other people's reactions to what you do or say?	.418	.085
Impulsivity9 Do you need to use a lot of self-control to keep out of trouble?	.393	.219
Impulsiveness10 I find long journeys boring – all I want is to get there as quickly as possible	.382	.188
Impulsiveness8 I seldom 'put my foot in it' *	.323	.041
Impulsiveness1 I seldom feel irritable *	.286	.184
Impulsivity 3 Do you like to arrive at appointments in plenty of time? *	.104	.094
Aggression5 If a passing car splashes me, I shout at the driver	.214	.465
Aggression7 I tend to snap at people	.434	.441
Aggression8 If someone says something stupid, I tell them so	.034	.427
Aggression10 I'd rather concede an issue than get into an argument *	-.083	.420
Aggression12 I don't think I could ever 'turn the other cheek'	.187	.368
Impulsiveness12 I hate being stuck behind a slow driver	.259	.279

** Items are reverse scored*

2. Examining the Discriminant and Predictive Validity of the Newly Devised Scales

Pre and Post-Treatment Scores and Effect Sizes

Summing the new scales within each SARN domain produced raw domain scores both pre- and post-treatment, with higher scores indicating greater problems within these three domains. Table 6.6 presents the pre and post domain scores for all those for whom reconviction data was available by reconviction group. The scores for the whole group on whom the scales were generated are also presented for reference. Finally, the table presents an effect size for each measure by group type using Cohen's *d*. These effect sizes represent the amount of pre to post change within-subjects by reconviction group. The n's vary for each measure as only those with both pre and post scores generated were included here.

The results indicate that improvements are made pre to post-treatment on all three domain measures and the total score for all subgroups. All generated effect sizes are medium or large, with the majority being large. Effect sizes are greatest for Total scale scores, followed by the Attitudes domain scale and smallest for the Self Management domain scale. In general, recidivists scored the worse on all three measures and the total score, with sexual recidivists scoring the worst on the Attitudes, Relationships and total scale scores, and sexual or violent recidivists scoring the worst on the Self Management Domain scale. Sexual recidivists produced the greatest effect sizes for all three measures and the total score, indicating that this group changed the most from pre to post-treatment on these scales. It should be noted however that this group of sexual recidivists is very small, and therefore caution is needed when interpreting these results.

Table 6.6: Raw Domain Scores and Cohen's *d* by Reconviction Status

		n	Pre-Treatment Mean (*SD*)	Post-Treatment Mean (*SD*)	*d* change
Attitudes Domain Score	Whole Sample	2376	28.21 (18.49)	13.94 (12.55)	1.10
	Sexual Recidivists	10	38.23 (23.04)	17.90 (15.72)	1.58
	Sexual Non-recidivists	878	28.07 (18.40)	13.89 (21.51)	0.84
	Sexual and/or Violent Recidivists	27	30.65 (21.01)	15.39 (15.26)	1.21
	Sexual and/or Violent Non-recidivists	861	28.13 (18.41)	13.89 (12.46)	1.09
	General Recidivists	80	29.41 (18.45)	15.10 (13.84)	1.09
	Non Recidivists	808	28.08 (18.50)	13.82 (12.41)	1.09
Socio-affective Functioning Domain Score	Whole Sample	1545	105.54 (26.68)	90.86 (24.10)	0.85
	Sexual Recidivists	7	123.88 (17.37)	102.89 (13.60)	1.85
	Sexual Non-recidivists	748	105.36 (26.70)	90.74 (24.16)	0.89
	Sexual and/or Violent Recidivists	18	120.41 (22.70)	103.67 (17.95)	0.99
	Sexual and/or Violent Non-recidivists	737	105.14 (26.68)	90.54 (24.16)	0.89
	General Recidivists	60	111.17 (24.44)	97.68 (23.65)	0.89
	Non Recidivists	695	105.02 (26.83)	90.28 (24.06)	0.89
Self Management Domain Score	Whole Sample	2397	14.94 (10.66)	9.73 (9.05)	0.58
	Sexual Recidivists	9	18.33 (11.34)	10.80 (7.84)	1.21
	Sexual Non-recidivists	888	14.90 (10.65)	9.71 (9.06)	0.58
	Sexual and/or Violent Recidivists	24	21.00 (12.08)	13.33 (11.05)	0.72
	Sexual and/or Violent Non-recidivists	873	14.76 (10.57)	9.62 (8.97)	0.58
	General Recidivists	79	20.49 (11.40)	14.09 (11.40)	0.59
	Non Recidivists	818	14.38 (10.37)	9.30 (8.68)	0.59
Total Score	Whole Sample	1622	148.82 (44.70)	115.06 (37.56)	1.14
	Sexual Recidivists	8	181.25 (38.91)	130.44 (31.87)	3.59
	Sexual Non-recidivists	787	147.84 (43.63)	113.17 (35.40)	1.22
	Sexual and/or Violent Recidivists	20	168.10 (40.07)	134.38 (36.07)	1.23
	Sexual and/or Violent Non-recidivists	775	147.67 (43.68)	112.80 (35.22)	1.23
	General Recidivists	68	159.07 (42.80)	126.42 (37.49)	1.15
	Non Recidivists	737	147.16 (43.66)	112.23 (35.00)	1.23

Examination of Pre and Post-Treatment Psychometric Scores by Recidivism Status

T-tests were conducted to compare the scores on each of the domains and the total score between those who were reconvicted of any reoffending and those who were not, those reconvicted of sexual reoffending and those who were not, and those who were reconvicted of sexual and/or violent reoffending and those who were not. Pre- and post-treatment, there were no significant differences when examining the attitudes domain score between general recidivists and non-recidivists (pre (t (969) = -.649, $p > .05$), and post (t (940) = -.907, $p > .05$)). However, there was a significant difference between sexual recidivists and sexual non-recidivists on their pre-treatment attitudes domain score (t (969) = -1.97, $p < .05$); sexual recidivists had significantly worse scores on this domain pre-treatment in comparison to sexual non-recidivists. There were no differences on this domain score post-treatment between sexual recidivists and non-recidivists (t (940) = -1.005, $p > .05$). There were no significant differences between sexual and/or violent recidivists and non-recidivists on this domain score pre- (t (969) = -.746, $p > .05$) or post-treatment (t (940) = -.624, $p > .05$).

Regarding socio-affective functioning psychometric test scores, at the pre-treatment stage there were no significant differences between general recidivists and non-recidivists (t (851) = -1.874, $p > .05$), but at the post-treatment stage general recidivists had significantly worse socio-affective functioning domain scores than non-recidivists (t (853) = -2.44, $p < .05$). Sexual recidivists also scored significantly higher on the socio-affective functioning domain score than sexual non-recidivists pre treatment (t (851) = -1.96, $p < .05$), but not post-treatment (t (853) = -1.506, $p > .05$). Sexual and/or violent

recidivists also had a significantly higher socio-affective functioning domain score than sexual and/or violent non recidivists both pre-treatment (t (851) = -2.66, $p < .01$), and post-treatment (t (853) = -2.47, $p < .05$).

Regarding self-management psychometric scores, general recidivists had significantly worse scores than non-recidivists pre-treatment (t (103.0) = -4.69, equal variances not assumed, $p < .001$), and post-treatment (t (94.77) = -3.79, equal variances not assumed, $p < .001$). There were no significant differences between sexual recidivists and sexual non-recidivists on the self-management domain scores either pre- (t (981) = -1.110, $p > .05$) or post-treatment (t (973) = -.377, $p > .05$). Sexual and/or violent recidivists, however, had a significantly higher (worse) self-management domain score than sexual and/or violent non-recidivists pre-treatment (t (981) = -3.07, $p < .01$), and post-treatment (t (973) = -2.10, $p < .05$).

Finally, when looking at total scores, recidivists scored significantly worse than non-recidivists for all groups, apart from post-treatment scores for sexual recidivists (who did not score significantly worse than sexual non-recidivists). The statistics were: general recidivists vs. non-recidivists pre-treatment (t (793) = -2.16, $p < .05$) and post-treatment (t (799) = -3.09, $p < .01$); sexual recidivists vs. non sexual recidivists pre-treatment (t (793) = -2.16, $p < .05$) and post-treatment (t (799) = -1.46, $p > .05$); sexual and/or violent recidivists vs. non sexual and/or violent recidivists pre-treatment (t (793) = -2.07, $p < .05$) and post-treatment (t (799) = -2.77, $p < .01$).

Predictive Accuracy of Pre and Post-Treatment Scale Scores using ROC Analyses and Cohen's *d*.

Predictive accuracy was examined for each pre and post-treatment scale score for sexual, sexual and violent, and general recidivism (see Table 6.7). A total score was also computed at pre and post-treatment and its predictive accuracy examined. The Attitude Domain Scale demonstrated weak predictive accuracy for all outcomes, both pre and post-treatment. The Relationships Domain Scale fared slightly better, and produced a large effect at the pre-treatment stage when predicting sexual reconviction, and moderate effects at the post-treatment stage when predicting sexual reconviction, and at the pre- and post-treatment stages when predicting sexual and/or violent reconviction. The Relationships Domain Scale demonstrated weak prediction for general recidivism. The Self Management Domain, on the other hand was a weak predictor of sexual reconviction, both pre and post-treatment, and of sexual and/or violent reconviction at the post-treatment stage, but was a moderate predictor of general reconviction both pre and post-treatment, and a moderate predictor of sexual and/or violent reconviction pre-treatment. The pre-treatment total score was a good predictor of sexual reconviction, a moderate predictor of sexual and/or violent reconviction, and a poor predictor of general reconviction. The post-treatment total score was a moderate predictor of both sexual reconviction and sexual and/or violent reconviction, and a poor predictor of general reconviction. It should be noted that the confidence intervals are relatively large for some of the AUC levels.

Table 6.7 Predictive Accuracy of Pre and Post-Treatment Domain Scale Scores for Sexual, Sexual and/or Violent, and General Recidivism

Measure	*n*	Sexual			Sexual and/or Violent			General		
		AUC	95%CI	*d*	AUC	95%CI	*d*	AUC	95%CI	*d*
Pre Treatment Attitudes Scale	971	.63	.47, .79	0.49	.52	.41, .62	0.13	.52	.46, .58	0.07
Post Treatment Attitudes Scale	942	.57	.37, .77	0.21	.51	.39, .62	0.11	.52	.45, .58	0.10
Pre Treatment Relationships Scale	853	.73	.59, .86	0.82	.68	.58, .79	0.62	.58	.51, .64	0.24
Post Treatment Relationships Scale	858	.68	.55, .80	0.62	.68	.59, .78	0.62	.60	.53, .66	0.31
Pre Treatment Self Management Scale	983	.59	.44, .74	0.31	.65	.55, .75	0.55	.65	.59, .71	0.56
Post Treatment Self Management Scale	975	.60	.46, .73	0.13	.62	.52, .72	0.37	.64	.57, .70	0.47
Pre Treatment Total	795	.73	.57, .89	0.81	.65	.53, .77	0.49	.59	.52, .66	0.28
Post Treatment Total	801	.66	.49, .81	0.51	.68	.58, .79	0.47	.62	.55, .70	0.39

Predictive Accuracy of Treatment Change Scores using ROC Analyses and Cohen's *d*.

Predictive accuracy was also examined for treatment change scores for sexual, sexual and/or violent and general recidivism (see Table 6.8). A total change score was also computed, and is presented too. Following the work of Beggs and Grace (2011) and Olver et al. (in press), table 6.8 presents the relationships between both the raw change scores and the standardized residual change scores (RCZs) for each scale and reconviction outcome. RCZs were computed by regressing the raw psychometric change score (post-treatment minus pre-treatment score) on the pre-treatment score for each scale. This residual was then standardized and the association between the standardized residual and outcome was examined. Thus the following was computed: actual change score – predicted change score = RCZ. This attempts to examine the change score whilst controlling for where someone started. As discussed in previous chapters this is important to examine since those who have higher pre-treatment scores are also the most likely to show the greatest raw change. Indeed, on examination of the correlations between the three pre-treatment domain scale scores and their respective change scores, positive relationships emerged across all measures (r = .47 to .73, p < .001). This indicates that worse scores on these measures pre-treatment were associated with greater amounts of change pre- to post-treatment. Use of RCZs has recently been suggested as one of the best methods to examine psychometric treatment change (Beggs & Grace, 2011). The results indicate that treatment change scores were poor predictors of all reconviction outcomes. The only exception to this was the total change score which was a moderate predictor of sexual reconviction.

Table 6.8 Predictive Accuracy of Change Scores for Sexual, Sexual and/or Violent, and General Recidivism

Measure	Sexual Reconviction					Sexual and/or Violent Reconviction					General Reconviction				
	n	AUC*	95%CI	*d*	*d* RCZ	*n*	AUC	95%CI	*d*	*d* RCZ	*n*	AUC	95%CI	*d*	*d* RCZ
Attitudes	888	.62	.48, .78	0.40	0.42	888	.50	.40, .60	0.02	0.03	888	.48	.41, .55	-0.05	-0.04
Relationships	755	.58	.41, .75	0.24	0.26	755	.48	.35, .61	-0.11	-0.09	755	.46	.39, .53	-0.17	-0.17
Self Management	897	.62	.47, .77	0.32	0.34	897	.61	.48, .74	0.39	0.43	897	.54	.46, .61	0.19	0.23
Total	668	.66	.54, .79	0.48	0.45	668	.54	.40, .68	0.02	0.00	668	.49	.41, .57	-0.10	-0.01

* The AUC figures are based on the raw treatment change scores. AUCs were not presented for the RCZs because there was little difference in the effect sizes between raw change scores and RCZs.

Cox Regression Analyses Examining the Predictive Accuracy of Pre, Post and Change Scores

A series of nested Cox Regression analyses were then performed to determine whether the psychometric domain scores could predict recidivism outcome. First, the sexual reconviction outcome was explored (see Table 6.9). As described in the data analytic plan, the pre-treatment scores were used as the baseline models, and the AICs were used to compare the fit of nested models. Table 6.9 indicates that none of the newly created Domain scale scores were significant predictors of sexual reconviction, and none of the baseline AIC values (pre-treatment scores) were improved upon. These findings could be due to the very low base rate of sexual recidivists in this sample ($N = 7$).

The next two tables present the results from the nested models conducted for the scales in predicting sexual and/or violent reconviction (Table 6.10) and general reconviction (Table 6.11). For sexual and/or violent reconviction, the results indicate that the Attitudes scale score is predictive of sexual and/or violent reconviction at the pre-treatment stage, but the post-treatment score and the change score are not predictive, and no models can improve the AIC produced from the baseline (pre-treatment) model.

Table 6.9 Summary of Nested Models in Prediction of Sexual Reconviction

		Hazard Ratio	95% CI	Chi-square	*p*-value	AIC
Attitudes Scale						
Baseline Model: Pre-Treatment Score		1.02	0.99-1.05	2.19	0.14	**122.59**
Post-Treatment Score		1.02	0.98-1.06	0.86	0.35	123.76
Pre- and Post-Treatment Scores	Pre	1.02	0.99-1.06	1.32	0.25	124.59
	Post	1.00	0.95-1.06	0.00	0.98	
Change Score		1.02	0.99-1.06	1.32	0.25	123.38
Pre-Treatment and Change Scores	Pre	1.02	0.98-1.06	0.89	0.35	124.59
	Change	1.00	0.95-1.06	0.00	0.98	
Relationships Scale						
Baseline Model: Pre-Treatment Score		1.01	0.99-1.03	0.28	0.60	**343.88**
Post-Treatment Score		1.01	0.98-1.04	0.59	0.44	343.60
Pre- and Post-Treatment Scores	Pre	1.00	0.98-1.03	0.00	0.97	345.60
	Post	1.01	0.97-1.05	0.29	0.59	
Change Score		1.00	0.97-1.03	0.00	0.99	344.15
Pre-Treatment and Change Scores	Pre	1.01	0.98-1.04	0.29	0.59	345.60
	Change	1.00	0.97-1.03	0.00	0.99	
Self Management Scale						
Baseline Model: Pre-Treatment Score		1.00	0.99-1.01	0.05	0.82	**1022.68**
Post-Treatment Score		1.01	0.99-1.03	1.02	0.31	1021.76
Pre- and Post-Treatment Scores	Pre	1.00	0.98-1.01	0.31	0.58	1023.44
	Post	1.01	0.99-1.04	1.27	0.26	
Change Score		0.99	0.98-1.01	0.36	0.55	1022.36
Pre-Treatment and Change Scores	Pre	1.01	0.99-1.03	0.97	0.32	1023.44
	Change	0.99	0.97-1.01	1.27	0.26	

Table 6.10 Summary of Nested Models in Prediction of Sexual and/or Violent Reconviction

		Hazard Ratio	95% CI	Chi-square	*p*-value	AIC
Attitudes Scale						
Baseline Model: Pre-Treatment Score		1.03	1.00-1.06	3.97	< 0.05*	**79.83**
Post-Treatment Score		1.03	1.00-1.06	3.17	0.07	80.82
Pre- and Post-Treatment Scores	Pre	1.02	0.98-1.07	1.14	0.29	81.73
	Post	1.01	0.96-1.05	0.10	0.75	
Change Score		1.01	0.97-1.06	0.35	0.55	83.42
Pre-Treatment and Change Scores	Pre	1.03	1.00-1.06	3.90	<0.05*	81.73
	Change	0.99	0.95-1.04	0.10	0.75	
Relationships Scale						
Baseline Model: Pre-Treatment Score		1.02	1.00-1.04	5.07	<0.05*	**216.89**
Post-Treatment Score		1.03	1.01-1.04	7.65	<0.01**	214.70
Pre- and Post-Treatment Scores	Pre	1.00	0.98-1.03	0.08	0.78	216.63
	Post	1.02	0.99-1.05	2.28	0.13	
Change Score		0.99	0.97-1.02	0.16	0.69	221.75
Pre-Treatment and Change Scores	Pre	1.03	1.01-1.05	7.37	<0.05*	216.63
	Change	0.98	0.95-1.01	2.28	0.13	
Self Management Scale						
Baseline Model: Pre-Treatment Score		1.01	1.00-1.02	2.20	0.13	741.18
Post-Treatment Score		1.01	1.00-1.02	5.60	<0.05*	**737.89**
Pre- and Post-Treatment Scores	Pre	1.00	0.98-1.01	0.30	0.58	739.59
	Post	1.02	1.00-1.03	3.62	0.06	
Change Score		0.99	0.98-1.00	1.13	0.29	742.22
Pre-Treatment and Change Scores	Pre	1.01	1.00-1.02	4.69	<0.05*	739.59
	Change	0.98	0.97-1.00	3.62	0.06	

* Significance at .05 level. ** Significance at .01 level.

Table 6.11 Summary of Nested Models in Prediction of General Reconviction

		Hazard Ratio	95% CI	Chi-square	*p*-value	AIC
Attitudes Scale						
Baseline Model: Pre-Treatment Score		1.04	0.98-1.10	1.40	0.24	**109.47**
Post-Treatment Score		1.02	0.95-1.09	0.27	0.60	110.54
Pre- and Post-Treatment Scores	Pre	10.4	0.97-1.11	1.19	0.28	111.43
	Post	0.99	0.91-1.07	0.04	0.84	
Change Score		1.03	0.96-1.10	0.82	0.36	110.03
Pre-Treatment and Change Scores	Pre	1.03	0.96-1.11	0.65	0.42	111.43
	Change	1.01	0.93-1.10	0.04	0.84	
Relationships Scale						
Baseline Model: Pre-Treatment Score		1.06	1.02-1.09	10.24	<0.001***	**293.50**
Post-Treatment Score		1.04	0.99-1.08	3.13	0.08	300.57
Pre- and Post-Treatment Scores	Pre	1.06	1.01-1.10	7.60	<0.01**	295.47
	Post	0.00	0.95-1.04	0.03	0.85	
Change Score		1.04	1.00-1.09	4.65	<0.05*	299.10
Pre-Treatment and Change Scores	Pre	1.06	1.01-1.10	6.36	<0.05*	295.47
	Change	1.00	0.96-1.05	0.03	0.85	
Self Management Scale						
Baseline Model: Pre-Treatment Score		1.05	1.03-1.07	26.91	<0.001***	**979.28**
Post-Treatment Score		1.05	1.02-1.07	19.52	<0.001***	987.93
Pre- and Post-Treatment Scores	Pre	1.04	1.02-1.07	11.12	<0.001***	979.35
	Post	1.02	0.99-1.04	1.95	0.16	
Change Score		1.02	0.99-1.05	3.44	0.06	1001.72
Pre-Treatment and Change Scores	Pre	1.06	1.04-1.08	27.53	<0.001***	979.35
	Change	0.98	0.96-1.01	1.95	0.16	

* Significance at .05 level. ** Significance at .01 level. *** Significance at .001 level.

The Relationships scale score is predictive at both the pre-treatment and the post-treatment stages, but the post-treatment model does not improve significantly upon the pre-treatment model (according to AIC values). Interestingly the model which includes both pre- and post-treatment scores does not produce any significant predictors; it appears that including both scores in the model dilutes any effect produced by either score. The model which includes both pre-treatment score and a change score is not an improved model to the pre-treatment score alone, suggesting that taking into account change does not add predictive power to the pre-treatment baseline score for the Relationships scale.

Finally, the Self Management scale score is a better (and significant) predictor at the post-treatment stage, and the AIC for this model is a significant improvement upon the baseline (pre-treatment score) model. This is an interesting difference to previous findings that the pre-treatment scores are generally the best predictors of recidivism outcome. Similar to the finding with the Relationships scale when including both pre- and post-treatment scores in the same model, neither of the Self Management predictors are significant. However, when including a change score in a model with the pre-treatment score, the latter becomes a significant predictor of sexual and/or violent reconviction. This indicates that taking into account change impacts on the effectiveness of the pre-treatment score of the Self Management scale when predicting sexual and/or violent reconviction.

When predicting general reconviction (Table 6.11), the results suggest that the Attitudes Scale score is not a significant predictor. The Relationships Scale score however

significantly predicts at the pre-treatment stage, and none of the additional models significantly improve upon the baseline (pre-treatment) model, according to the AIC values. The Relationship Scale change score is also a significant predictor of general reconviction, but when included in a model with the pre-treatment score it does not remain a significant predictor. This suggests that taking into account the change score does not impact, or add predictive power to, the pre-treatment Relationships Scale score.

Finally, the Self Management scale score significantly predicts general reconviction at the pre-treatment stage, and the post-treatment stage. However, the post-treatment model is not an improvement on the baseline (pre-treatment) model, suggesting that the pre-treatment score is the better predictor of general reconviction of the two. When in a model together, the post-treatment score does not add to the predictive power of the pre-treatment score, and similarly when the change score is added to a model with the pre-treatment score it does not add to the predictive power of this baseline score.

Incremental Validity

A final set of Cox Regressions were conducted to examine the incremental validity of the pre and post-treatment, and total scores alongside static risk for the three different outcomes (general, sexual and sexual and/or violence). The static risk assessment used was dependent on the type of reconviction being predicted; for general recidivism the two-year OGRS score was used, whereas for sexual recidivism, and sexual and violent recidivism (in the absence of a specific predictor of sexual and violent recidivism), the RM2000/s score was used.

The first three analyses examined the incremental validity of the three pre-treatment scale scores for the three outcomes separately. None of the three pre-treatment scores could add to the predictive power of the OGRS tool when predicting general reconviction. The OGRS tool however was highly predictive (Wald = 82.04, $p < .001$). Equally, none of the pre-treatment scores were significant predictors in the model for sexual reconviction. Similarly to findings from chapter two, the RM2000/s was not a significant predictor of sexual reconviction in this analysis, due to the very low base rate of sexual reconviction. The results of these two analyses are shown in Appendix E. When predicting sexual and/or violent reconviction, the pre-treatment Self Management Scale score significantly added to the predictive power of the RM2000/s, as shown in Table 6.12. Interestingly, this score reduced the predictive power of RM2000/s in step 2 of this model.

Table 6.12 Summary of Cox Regression examining Incremental Validity of Pre-Treatment Scores predicting Sexual and/or Violent Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: Sexual and/or Violent Recidivism					
Step 1: RM2000	0.57	0.27	4.33	<0.05	1.76
Step 2: RM2000	0.25	0.29	0.72	0.40	1.28
Pre-Treatment Attitudes Score	-0.01	0.02	0.60	0.44	0.99
Pre-Treatment Relationships Score	0.01	0.01	1.09	0.30	1.01
Pre-Treatment Self Management Score	0.05	0.03	4.06	<0.05	1.05

The next three regression analyses examined the incremental validity of the three post-treatment scale scores for the three outcomes separately. The findings were very similar to those described above for the pre-treatment scale scores. None of the post-treatment scores could add to the predictive power of the OGRS tool when predicting general reconviction (but again this was highly significant in the model; Wald = 70.96, $p < .001$), or the RM2000/s when predicting sexual and/or violent reconviction (RM2000: Wald = 6.91, $p < .01$). None of the post-treatment scale scores were predictive of sexual reconviction, and the RM2000/s was also not predictive of this outcome. (Results of analyses are presented in Appendix E).

The final three regression analyses examined the incremental validity of the total scale scores for the three separate outcomes (general, sexual and sexual and/or violent reconviction). For these analyses a total pre and post-treatment score was used based upon the sum of the three domain scores. The model predicting general recidivism was significant (*-2LL* = 589.122, χ^2 = 6.29, $p < .01$), with both static risk (Wald = 58.27, $p < .001$), and the post-treatment total score (Wald = 5.60, $p < .05$) being significant predictors of outcome (see Table 6.13). The second model predicting sexual reconviction was not significant (*-2LL* = 51.297, χ^2 = 3.17, $p > .05$) and none of the predictors were significant. Within the third model, predicting sexual and/or violent reconviction, the model was significant (*-2LL* = 152.092, χ^2 = 5.94, $p < .01$), and RM2000/s was a significant predictor (Wald = 4.02, $p < .05$), but neither the pre nor the post total score added incremental validity to this. Analyses for the two latter models are shown in Appendix E.

Table 6.13 Summary of Cox Regression examining Incremental Validity of Pre-Treatment and Post-Treatment Total Scores predicting General Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: General Recidivism					
Step 1: OGRS	4.63	0.59	61.52	<0.001	102.81
Step 2: OGRS	4.69	0.62	58.27	<0.001	109.06
Pre-Treatment Total Score	-0.01	0.01	1.55	0.21	0.99
Post-Treatment Total Score	0.01	0.01	5.60	<0.05	1.01

Chapter Discussion

This chapter attempted to develop and validate a short psychometric test battery using a larger, existing battery, to measure three dynamic risk domains associated with sexual reoffending; offence-supportive attitudes, socio-affective functioning and self-management.

The first stage was to eliminate any measures which were not deemed to be useful in the production of new scales according to a set of predefined criteria. These criteria excluded any measures which met the following conditions:

1) do not directly measure the treatment targets of the sex offender treatment programmes in custody;
2) are copyright and cannot therefore be changed/altered in any way (as such cannot be used in the generation of new scales);
3) have been found to be unrelated to recidivism in previous research;
4) were removed from the NOMS psychometric battery since the 2003 revisions.

This process in itself reduced the item pool of the battery from 457 to 192. The second stage involved using factor analyses on the items in the remaining battery to identify potential items to retain, and ultimately to determine whether risk domain, rather than individual risk factor, scales could be produced with adequate psychometric properties.

The factor analytic process reduced the total item pool from 192 to a total item pool of 147, which equates to a 24% reduction in the number of items. Although this reduction

was not as large as hoped, it did produce meaningful factors. Factor analyses revealed a three factor solution was preferred for items pertaining to child abuse supportive beliefs. These factors were entitled denial of harm, child as sexual, and normalising sexual contact with children. A one factor solution was preferred for the items pertaining to negative attitudes towards women. All generated scales in the attitudes domain had good to excellent internal consistency, as did the overall attitudes domain score produced by adding together the scores from the four separate scales and which was used in further analyses. Regarding the socio-affective functioning items, a number of factors emerged from the factor analyses, suggesting that the different measures in this domain are measuring a number of discreet, albeit theoretically related, constructs. The internal consistency of these scales ranged from moderate to excellent, and the overall socio-affective functioning scale had excellent internal reliability. Factor analysis of the scales in the self-management domain resulted in a two-factor solution, measuring impulsivity and aggression respectively. The impulsivity scale had an excellent internal consistency, and the aggression scale had a moderate internal consistency.

Scores of the newly created domain scales were then examined. Significant and positive changes pre- to post-treatment were observed for the whole sample and all recidivist/non-recidivist groups for the three domain scales and the overall psychometric score. The biggest changes were observed on the total psychometric score and the Attitudes scale, and the smallest for the Self Management scale. These findings suggest that the domain scales are able to detect treatment change.

On examination of the discriminant validity of the scales, the Attitudes scale could discriminate between sexual recidivists and sexual non-recidivists pre-treatment, with the recidivists scoring worse. On the Socio-affective functioning scale general recidivists scored worse than non-recidivists post-treatment, sexual recidivists scored worse than non sexual recidivists pre-treatment and sexual and/or violent recidivists scored worse than non-recidivists both pre- and post-treatment. On the Self Management scale both general recidivists, and sexual and/or violent recidivists scored worse than non-recidivists pre- and post-treatment. On the total scale, recidivists scored significantly worse than non-recidivists for all groups, apart from post-treatment sexual recidivists (who did not score significantly worse than sexual non-recidivists). These results show some indication that the scale scores can discriminate between recidivists and non-recidivists.

The predictive validity of the domain scales were examined. AUC and Cohen's *d* analyses found the scales generally to have weak to moderate predictive ability (summarised in Table 6.14). The Attitudes scale was a weak predictor of all outcomes. The Relationships scale was a moderate predictor of sexual and sexual and/or violent recidivism, but weak predictor of general recidivism. The Self Management scale was a moderate predictor of general recidivism, but a weak predictor of sexual (and sexual and/or violent) recidivism. The overall scores were moderate to good predictors of sexual recidivism and sexual and/or violent recidivism, and poor predictors of general recidivism. The change scores were poor predictors of all reconviction outcomes (when using both raw change scores and RCZ scores). However, the total change score was a moderate predictor of sexual reconviction (demonstrated when using both types of

change scores). It should be noted that for some of the AUCs, the confidence intervals were relatively large, indicating that less confidence can be placed on the associated effect sizes.

Table 6.14 Summary of each of the Scales' Level of Prediction of Different Reconviction Outcomes

	Outcome		
	General Recidivism	Sexual Recidivism	Sexual and/or Violent Recidivism
Attitudes Scale	Weak	Weak	Weak
Relationships Scale	Weak	Moderate	Moderate
Self Management	Moderate	Weak	Weak
Total Score	Weak	Moderate	Good
Total Change Score	Weak	Moderate	Weak

Cox regression analyses similarly found that scale scores were only moderately predictive of recidivism outcome. None of the models predicting sexual recidivism were significant for any of the scales. This is most likely due to the low recidivism rates of the sample. Overall, pre-treatment scores were generally the best in predicting sexual and/or violent recidivism and general recidivism, and adding post-treatment scores, or change scores to the pre-treatment scores did not, in general, add to the predictive power of these scores. There were some exceptions to this; the post-treatment Self Management scale score was a better predictor of sexual and/or violent recidivism than the pre-treatment score, and the change score also added to the predictive ability of the pre-treatment score on this scale. Additionally the post-treatment Socio-affective Functioning scale scores were equally

predictive of sexual and/or violent recidivism as the pre-treatment scores. The current findings in general therefore support previous research which has mostly found that it is pre-treatment rather than post-treatment scores which have a stronger relationship with reconviction, possibly as there is less incentive to respond in a socially desirable way at the pre-treatment stage, when starting treatment, than at the post-treatment stage when scores are used to inform risk assessment and therefore decisions about further treatment or release (Barnett et al., 2012, 2013; Wakeling et al., 2011).

When examining incremental validity, the domain scale scores were generally shown to be weaker predictors of recidivism outcome than static risk assessment tools. However, the pre-treatment Self Management scale score was a significant predictor when added to a model with RM2000/s when predicting sexual and/or violent recidivism. And the total post-treatment psychometric total scale score added to the predictive ability of the OGRS tool when predicting general recidivism. These results indicate that an overall psychometric score based on the post-treatment psychometrics generated in the present study may be a useful indicator of which offenders may be more likely to go on to be reconvicted of a further offence. Wakeling et al. (2011) also found that a measure of overall post-treatment dysfunction using amalgamated psychometric test scores was predictive of recidivism, although this study examined sexual and/or violent recidivism alone, and not general reoffending as well, as in the present study.

The Attitudes scale in general was the weakest performing of the three scales. This is in line with previous research on psychometric tests which has found that amalgamated

measures of offence-supportive attitudes, pre- or post-treatment, did not predict reconviction outcome (Barnett et al., 2012, 2013; Nunes et al., 2014; Wakeling et al., 2011). Such measures, which have clear demand characteristics, may be more likely to be subject to socially desirable responding, decreasing the reliability of the results of these tests. The Relationships and Self Management domain scale scores fared slightly better. In particular, the Self Management scale appeared to be relatively good at predicting general recidivism, and the Relationships scale appeared to be relatively good at predicting sexual and/or violent recidivism.

The current research did to some extent achieve its aims. Firstly the study was successful to an extent at reducing the psychometric battery to a smaller number of items, whilst retaining meaningful constructs of relevance to sexual offenders. Although not as large a reduction as anticipated, the revised psychometrics are still slightly shorter in length, and would therefore both reduce the amount of resources required to administer them, and the length of time offenders would need to complete them. Whilst there is undoubtedly disadvantages to producing new scales instead of utilising scales in their current validated forms (the need for further validation just one consequence of doing this), the drive for efficiency within rehabilitation services in the UK propels us into attempting to minimise resource implications for effective practice. However, further research is required before implementation of a shortened version. For example, it would be useful to examine the extent to which the full psychometric battery version and the shorter version produced in the current research are able to identify treatment needs or clinically significant problems, and whether in practice one is preferable to the other in performing this function.

Limitations

There are a number of limitations relevant to the current study. As mentioned, the very low rate of sexual recidivism in the sample, may have reduced the power to detect effects when this outcome was used in analyses. A longer follow-up or larger sample would help to determine whether this was the case. In addition, the sample comprised only of those who have been convicted and incarcerated of a sexual offence and had undertaken, and indeed, completed, treatment for sexual offending. The sample is therefore not representative of sexual offenders more generally, and the exclusion of those refusing or dropping out of treatment is likely to have led to quite different results for the analyses of the pre-treatment psychometric test scores, than had these groups been part of the sample. Finally, the Sexual Interests Domain psychometrics were not examined within the present study.

In conclusion, the results do suggest that a short battery of tests designed to measure problems in broad areas related to sexual reoffending, can be both reliable and reasonably valid indicators of these constructs. The findings suggest that the measures of socio-affective functioning and self-management can provide minimally useful information about risk of general recidivism as well as sexual and/or violent recidivism to an extent, perhaps measuring the more criminogenic needs in sexual offenders. The psychometric tests appear less able to provide good measures of dynamic risk factors associated specifically with sexual recidivism, although further research is required before any firm conclusions can be drawn.

CHAPTER SEVEN

DISCUSSION AND CONCLUSIONS

This thesis aimed to examine the use of self-report psychometric measures within Sex Offender Treatment Programmes and particularly their ability to predict reconviction. This overarching aim has been achieved by setting out how psychometrics are currently used within treatment programmes and summarising previous research (chapters 1 and 5); by conducting empirical studies examining the predictive validity of the pre and post-treatment psychometrics (chapter 2) and which ones are the best predictors in comparison with static factors (chapter 3); by conducting an empirical study examining the predictive validity of treatment change as measured psychometrically (chapter 4); and by attempting to produce a smaller, more refined psychometric battery with good predictive validity (chapter 6). The aim of this final chapter is to bring all these findings together and provide some general conclusions and directions for further research. The implications of the findings for practice and policy will also be presented.

I will now present a summary of the conclusions by chapter, before addressing the five specific hypotheses which were listed in chapter 1.

Summary of Conclusions by Chapter

Chapter 1 provided a thorough introduction and literature review of the use and utility of psychometrics within the Sex Offender Treatment Programmes, and previous research which has examined the discriminant and predictive validity of psychometrics within the forensic field. This chapter therefore attempted to set the picture for the remaining chapters of the thesis. The conclusions from the literature review indicated that psychometric measures are often used within treatment and evaluation contexts to help identify treatment need, to examine individual treatment change, to contribute to risk assessment, and to help evaluate programme's effectiveness. However, the research literature is not clear as to whether psychometrics have discriminant and predictive validity; some studies have found that they do, whereas others have not. To be used as reliable indicators of risk, need and change, psychometrics must be shown to have good validity. The research findings seem to depend on how the psychometrics have been gathered and used, which psychometrics have been used, how the relationship between the tests and reconviction has been examined statistically, and the samples used. The review provided a clear need for further examination of the discriminant and predictive validity of a series of psychometrics which are routinely used on the NOMS Sex Offender Treatment Programme, and to determine how they should be used within the treatment of sexual offenders.

Chapter 2 explored the relationship between individual psychometric tests at pre-treatment and post-treatment time points, and sexual and/or violent reconviction. A large number of psychometric tests (used at various points on the NOMS Sex Offender

Treatment Programme) between 1996 and 2006 were examined. The results found that a number of psychometric tests could discriminate recidivists from non-recidivists. The individual psychometrics which had some discriminant validity were mainly those assessing socio-affective functioning and self management problems. When scores were clustered into SARN domain scores, they continued to show discriminant validity, particularly at the pre-treatment stage. Predictive validity was examined using regression analyses, which found that the socio-affective functioning domain score was a significant predictor of sexual and/or violent recidivism, and added incremental validity to a measure of static risk. An overall level of dysfunction based on the psychometrics also had good discriminant validity, and added to the predictive power of static risk at the pre-treatment stage. The findings from chapter 2 therefore indicate that psychometrics may have some predictive validity, particularly at the pre-treatment stage and for those measures assessing socio-affective functioning, and there appears to be some usefulness in clustering psychometric scores into risk domain scores.

Chapter 3 used prognostic modelling techniques to further examine the ability of the pre and post-treatment psychometric scores to predict recidivism outcome. Notably, this study attempted to identify not whether each psychometric could predict outcome or distinguish between recidivist groups (as chapter 2 did), but which of the available variables (both static and psychometric/dynamic) were the best predictors of reconviction when entered into a prognostic model together. The findings indicated that the best predictors of sexual and/or violent reconviction were age at release from prison, number of sexual appearances, and number of criminal appearances. Thus, in the present study,

individual static factors were better predictors of outcome than any of the individual psychometric variables. The pre-treatment impulsivity score was the best predictor of the psychometric variables used in the present study, and similar to findings from chapter 2, the pre-treatment socio-affective functioning domain score (when clustering all psychometrics from this domain together) was predictive of recidivism in a separate model, as was a pre-treatment overall psychometric score of dysfunction. Absolute risk estimations were also produced within this chapter based on the final significant prognostic model providing a potential way of estimating survival rates for offenders based on their characteristics of the three static predictors in the model (see appendix C).

Chapter 4 examines the relationship between psychometric change and reconviction using clinical significant change methodology. The findings indicate that those who scored within a 'normal' range before and after treatment had lower reconviction rates than those who scored 'worse than normal' after treatment on a selection of psychometrics. These psychometrics were mainly from the socio-affective functioning domain but additionally this finding emerged for the sexual obsessions subscale of the MSI and a measure of impulsivity. This suggests that these psychometric scores appear to have some relationship with recidivism group, and that those who score in a normal range are less likely to go on to be reconvicted of a sexual or violent offence. Additionally, those who were deemed to have 'changed' psychometrically (by domain) were less likely to be reconvicted than those who did not change. This was found for the sexual interests domain, the socio-affective functioning domain and the self management domain. An overall treatment change status was also found to discriminate between

recidivists, and was a modest predictor of sexual and/or violent reconviction (though did not add incremental validity to static risk). This study provided some preliminary support for the use of clinical significant change methodology in identifying who might be more likely to go on to be reconvicted of a further offence, and could potentially be used as a clinical tool on an individual basis. Again the study indicated that measures from the socio-affective functioning domain might be the most discriminatory and predictive, that pre-treatment scores may be slightly better predictors than post-treatment scores, and that domain change scores can be useful.

Chapter 5 presented a summary of the findings of the three empirical studies conducted thus far within the thesis and presented these alongside other emergent literature in the field. The main conclusions were that the research on the NOMS Sex Offender Treatment Programme psychometrics has found that they are only moderately predictive of future reconviction. Specifically the following conclusions were drawn: a) clustering psychometrics into risk domains seems to enhance their predictive power; b) measures examining socio-affective functioning seem to have the greatest predictive validity; c) an overall psychometric deviancy score may enhance predictive validity; d) pre-treatment psychometrics are generally better predictors than post-treatment psychometrics; e) psychometric change scores are associated with recidivism for a small selection of measures; and f) psychometric scores are generally less predictive than static risk variables. Other emergent literature since this thesis had started provided further indication that the methodology used to assess treatment change may have an impact on the predictive validity findings within empirical studies as may the context in which

psychometric measures are used. Taken together the main conclusion from this review chapter was the need to attempt to develop a new psychometric battery which specifically measures risk domains (rather than individual risk factors), and to examine the relationship between psychometric scores and change and reconviction using different methodology.

Chapter 6 attempted to produce a new, smaller psychometric battery based on SARN risk domains via theoretical and factor analytical techniques. A smaller battery was produced, which was found to have good psychometric properties, was able to detect treatment change, and had good discriminant validity. The domain scales (pre scores, post scores and change scores), however, only had weak to moderate predictive validity when examining three separate outcomes (sexual reconviction, sexual and/or violent reconviction, and general reconviction). This was observed by production of AUCs, Cohen's *d* and regression analyses. Thus, the new scales had similar predictive validity to the previous longer versions. Pre-treatment domain scores were generally better than post-treatment domain scores, or change scores at predicting outcome. The Socio-affective functioning domain scale score was generally a better predictor of sexual and violent reconviction, whereas the Self Management domain scale score was a better predictor of general reconviction. There were no differences in predictive validity of change scores dependent on the way these were calculated (raw or use of RCZ). In general, the new domain scales were unable to add incremental validity to static risk. This chapter advanced on previous methodology by examining three separate

reconviction outcomes, by using different methodology to determine change scores, and by the use of comparing nested models.

Hypotheses and Overall Findings

Linking these findings back to the overarching hypotheses set out in Chapter 1 can provide a useful way of summarising the main findings from this thesis.

Hypothesis 1: Pre- and post-treatment psychometrics will demonstrate good discriminant validity

A selection of psychometrics examined within this thesis demonstrated discriminant validity; that is scores on some measures were able to significantly distinguish between recidivists and non-recidivists. Most notably, a selection of measures from the socio-affective functioning domain appeared to show discriminant validity with recidivists scoring significantly worse than non-recidivists including: Entitlement to sex, ruminations, self esteem, empathic concern, perspective taking and locus of control. Further measures in the Self Management domain including a measure of impulsivity also demonstrated discriminant validity. When clustered into domains, all four SARN domains were able to discriminate recidivists from non-recidivists pre-treatment, but only the socio-affective functioning domain and the self management domain showed discriminant validity post-treatment. This hypothesis is partially supported.

Hypothesis 2: Pre- and post-treatment psychometrics will demonstrate good predictive validity

A small selection of psychometrics also demonstrated a moderate level of predictive validity, particularly the impulsivity scale and the MSI sexual obsessions scale. Predictive validity increased when psychometrics were clustered into risk domains. Most notably, an overall socio-affective functioning domain score appeared to be a moderately good predictor of sexual and/or violent reconviction. This hypothesis is partially supported.

The findings for both Hypotheses 1 and 2 indicate that those risk factors contained within the socio-affective functioning domain appear to be better at discriminating between those who go on to be reconvicted and those who do not, and better at predicting outcome, in comparison to the risk factors within the other three SARN domains. Certainly there is evidence that these risk factors, including inadequacy, a tendency to ruminate, and lacking relationships with adults, have a reliable relationship with recidivism (e.g., Hanson & Morton-Bourgon, 2005; Mann et al., 2010). However, there is also good evidence in the literature that all of the other risk factors within the SARN framework have a reliable relationship with recidivism too. As such, there are two equally plausible explanations for the present findings. First, it could be that the measures used within this domain are somehow better at measuring these risk factors than the psychometrics used to measure the risk factors within the other three SARN domains. For example, they may have better construct validity. Second, it could be that treatment participants may be more likely to respond in an honest and accurate way to self-report measures within this domain as the constructs therein could be construed as less intrusive and shame-provoking than the measures within, for example, the offence supportive

attitudes domains. It may be more likely that treatment participants respond in a socially desirable way to measures within the latter two SARN domains which are more offence-specific in content and perhaps more easy to fake (in that respondents are more likely to understand how they should answer to appear less deviant or more treated). As such, it may be the context in which these psychometrics are gathered and used which has an impact on their discriminant and predictive validity.

There is also indication that clustering psychometrics into domain scores could enhance the predictive power of the psychometrics. This could provide a useful and less resource intensive way of using self-report psychometrics in the treatment and risk assessment of sexual offenders. This will be discussed further in the implications for practice and policy section.

Hypothesis 3: There will be no difference in discriminant and predictive validity between pre- and post-treatment psychometrics

In general, a consistent finding throughout this thesis was that the pre-treatment psychometric scores (both individually and clustered into domains) were better discriminators between recidivism groups and better predictors of reconviction outcome than post-treatment psychometric scores. This hypothesis is therefore not supported.

These findings are in line with previous research (e.g., Hanson & Wallace-Capretta, 2000; Quinsey et al., 1998), though do not add support to the use of post-treatment psychometrics in contributing to an assessment of progress in treatment. The reason for

this finding could be to do with the context in which these psychometrics are being administered and the fact that it is known that they are being used to examine progress in treatment. As such, treatment participants could be more likely to respond in a favourable way at the post-treatment stage to appear more 'treated'. Being prone to faking and their transparency are indeed some of the limitations of self-report psychometrics, and perhaps because of this their use within the criminal justice system as measurements of change is questionable. As discussed in chapters 2, 3 and 4 this may also be to do with the fact that the post-treatment scores will contain a mix of individuals who have changed to varying degrees as a result of treatment (or a result of faking good) and therefore their relationship with outcome is less likely to be linear. Certainly, these results suggest that the post-treatment psychometric scores should be relied on less than the pre-treatment scores by those involved in the treatment and risk assessment of sexual offenders, as they have a less reliable relationship with reconviction outcome.

Hypothesis 4: Pre and post-treatment psychometrics will be as good as static factors in predicting reconviction

This hypothesis was not supported. Overall this thesis found that static factors (both individually as shown in chapter 3 and as a static risk level as shown in chapters 2, 4 and 6) were better predictors of reconviction outcome than psychometric scores. This finding supports the literature on the use and validity of actuarial static risk assessment tools (Hanson & Morton-Bourgon, 2009) as better predictive tools than unstructured clinical judgement and some structured clinical judgement tools. However it does not support other research which has found that psychometric variables can add to the predictive

ability of static tools (e.g.. Allan et al., 2007). Since this was a consistent finding across different chapters of this thesis, and most notably in chapter 3 with the use of prognostic modelling, it is a finding which is worthy of significant comment. However, these results do not mean that dynamic risk factors are not predictive of reconviction, that dynamic risk factors are less predictive than static factors, or that they are not worthy of examination within treatment. Contrary, there is overwhelming evidence that the risk factors within SARN are robust, and have a reliable relationship with reoffending (Mann et al., 2010), and that they can be as important as static variables (Thornton, 2002) in identifying risk of recidivism. The findings instead suggest that it is the self-report psychometric measures which are not entirely reliable measures of these dynamic risk factors. It is the dynamic risk factors *as measured by these self-report psychometrics* which are not as predictive as static items, not the dynamic risk factors themselves.

Hypothesis 5: Psychometric change scores will demonstrate good predictive validity

Arguably it is the treatment change scores which are fundamental to understanding the effect of treatment on recidivism outcome, and in identifying individuals who might be more likely to go on to recidivate. Overall, the findings in this thesis provided only some limited evidence that psychometric change can predict recidivism outcome, and change scores were, in general, less predictive than static risk scores, and less predictive than pre-treatment psychometric scores. These results were consistent when using various different methodologies to calculate change scores (clinical significant change in chapter 4 and raw change and RCZ scores in chapter 6). Therefore hypothesis 5 is not supported. These findings are consistent with other recent research (e.g. Barnett et al., 2013; Beggs

& Grace, 2011; Nunes et al., 2014), and suggest that the use of self-report psychometrics alone in examining individual treatment change may be inadequate.

Hypothesis 6: A smaller psychometric battery with equivalent or better discriminant and predictive validity can be created

This hypothesis was partially supported. A smaller psychometric battery based on SARN domains via factor analytical techniques was developed, which had good psychometric properties, demonstrated pre to post-treatment change, and could discriminate between recidivists and non recidivists (chapter six). However, the new scales were only weak to moderate predictors of recidivism outcome, and did not (in general) add incremental validity to static risk. Thus, although a smaller battery was created, its predictive validity was not an improvement on the original psychometrics examined in chapters 2-5. While savings in resource and time would be made by using this revised domain battery, there would be no advantages in terms of outcome prediction.

Overall Conclusions

To conclude, the present thesis has contributed to the debate on whether self-report psychometric variables can be useful in predicting recidivism. The most obvious conclusion which could be drawn from this thesis is that there is little support for the continued use of a large battery of psychometrics within sex offender treatment and risk assessment which does not facilitate identification of who might be more likely to go on to reoffend, and which does not appear to measure the known dynamic risk factors of recidivism particularly reliably. A more moderate conclusion would be that the self-

report psychometrics may in fact be useful predictors of recidivism, with small effect sizes by conventional standards (and not as large effects as static variables), but which may be meaningful in the sexual offender recidivism field, which is confounded with low base rates of recidivism and recidivism error measurement (e.g. Hanson & Morton-Bourgon, 2005; Helmus et al., 2013; Mann et al., 2010). As such, in line with Nunes et al. (2014), it is suggested that researchers and clinicians continue using psychometrics and examine them with samples that have longer and larger recidivism rates to examine their usefulness further, but to be less reliant on them in terms of identifying change, and contributing towards treatment progress estimates and risk assessment procedures.

The current findings are generally not consistent with the view that the psychometric measures reliably assess dynamic risk factors for sexual recidivism. A good measure of a dynamic risk factor has to be predictive of recidivism, and changes on that factor must be related to changes in recidivism risk (Andrews & Bonta, 2010; Douglas & Skeem, 2005; Harris & Rice, 2003; Seto, 2008). Whilst some of the measures examined in this thesis show some promise (and domain scores show greater promise), generally improvement in the measures was not shown to be related to recidivism outcome. As mentioned previously, the findings do not however indicate that the dynamic risk factors measured within SARN are not predictive. It is more likely that the measures themselves are not sensitive enough to measure these risk factors or that measuring these risk factors using psychometric measures is not the best method in identifying these risk factors. Similarly, the fact that psychometric change scores only had limited predictive validity does not indicate that offenders are not making changes in dynamic risk factors as a result of

treatment, rather that the psychometric measures may not be picking up these changes appropriately. This thesis, therefore, does not negate the importance of targeting the psychologically meaningful risk factors as per Mann et al. (2010) or the treatment needs within the SARN framework. The evidence shows us that these treatment need areas are robust and significant predictors of sexual recidivism. An alternative explanation, which should be considered as well, is that the psychometrics do measure the likelihood of reconviction, but that reconviction itself is not an accurate measure of reoffending. Reconviction might not reflect the true level of reoffending (due to issues of crime reporting and levels of detection), and this might affect the relationship between the variables and output examined.

Implications for Practice and Policy

These findings have important implications for clinicians working with sexual offenders and for policy leads in the design and development of treatment programmes and risk procedures for sexual offenders. Firstly, it would be wise for clinicians not to rely too heavily on the use of self-report psychometrics in identifying treatment need, treatment change and progress in treatment. Equally, it would be unwise for policy leads to place too much emphasis on the use of psychometrics as a measure of need, risk and change. As outlined in chapter one, psychometrics are currently used as one part of a triangulation of evidence for identifying treatment need and progress in treatment. This thesis supports a triangulation of evidence approach. On the basis of the present findings, it would be unethical and inadvisable to use self-report psychometrics as the sole basis for identification of dynamic risk factors and identifying treatment change. Placing less

emphasis on them, and using them in conjunction with other evidence, including observation, interview, and file information, seems a sensible strategy. Self-report psychometrics have a number of limitations, and this thesis has further found them to have only limited predictive validity.

Secondly, however, there is some indication that the psychometric measures used to identify risk factors within the socio-affective functioning domain may have better predictive validity than those measures identifying risk factors within the other SARN domains. Clinicians could thus perhaps be more confident in the use of these particular psychometrics when using them to determine risk and treatment change. Third, the findings provide some limited promise for the benefits of using risk domain scores (based on psychometrics) rather than individual psychometric scores, as generally these fared better in terms of predictive validity. Thus, clinicians should consider using clustered psychometric scores (or even the shortened battery of domain scores as generated in chapter 6) as a means of helping to identify who might be more likely to go on to reoffend or who might benefit from further treatment before release from prison. In this way, the clustered psychometrics could aid in the allocation of limited resources to those offenders most at risk of reoffending. Policy leads in particular should bear this in mind when developing future treatment programmes and procedures. Fourth, the findings suggest that static factors may be better and more reliable predictors of future reconviction than dynamic factors (as measured psychometrically). Those working in the treatment and risk assessment of sexual offenders should bear this in mind when formulating treatment plans and writing risk assessment reports, and ultimately when

making recommendations for release/risk management. As discussed previously, there is no indication from this research that dynamic risk factors are not important, rather it is suggested that psychometric measures may not be the best means of measuring these risk factors, and when psychometrics are used to measure these dynamic risk factors alone, that they are not as predictive as static factors.

The findings from this thesis do provide some potential promising avenues for those working with sexual offenders. Firstly, there is promise in the use of pre-treatment scores to identify those who might be less likely to need treatment (or to reoffend). For example, using the clinical significant change methodology (chapter four), the results indicated that offenders whose psychometric scores are *already ok* prior to treatment have the lowest likelihood of future reoffending. As such, these pre-treatment scores, in conjunction with a review of static risk, could help clinicians make decisions on which offenders should be offered treatment. In times when resources are often scarce, this could help prioritise sexual offender treatment places for those most in need of these services, and policy colleagues would be well advised to examine the use of psychometrics in this way for the future. It could be, for example, that those offenders who were *already* okay on the psychometrics might be more likely to benefit from other types of interventions, perhaps less offence-specific focused programmes which address general criminal attitudes or target thinking skills. Furthermore in the same study it was found that those who improve during treatment but not enough to make them appear *functional* post-treatment are perhaps those most likely to go on to reoffend. Having this information on individuals could aid in decisions regarding future treatment need and

release plans for those who have been on a treatment programme. From these findings, I suggest, similarly to Nunes et al. (2011), that psychometrics may be more useful to examine on an individual basis and to use as a clinical tool, rather than on a group basis or used as a research tool to examine the effectiveness of programmes. It is particularly important to note that research which examines the issue of dynamic risk factors using psychometrics alone may not be wholly representative. It is critical that those working in this field understand the distinction between the predictive power of dynamic risk factors and the predictive power of psychometric measures attempting to measure these dynamic risk factors. This is particularly important when reading or synthesising the literature on this topic and when making decisions about the relevance of different risk factors in the prediction of future behaviour.

Chapter three also provided information regarding the potential use of absolute risk estimations for individuals (using static factors). This could be a useful tool for clinicians, alongside typical risk assessment tools, as a means of providing estimations of survival rates for individuals based on their characteristics on a small number of items. Although these estimations were based on static factors, and not dynamic/psychometric factors (the key topic of this thesis), it is still a useful conclusion to be highlighted from the findings. Thus, there are some key findings which could have an impact on helping to identify those sexual offenders most in need of treatment, which could be extremely useful for clinicians and treatment providers.

Limitations

There are a number of limitations to the studies within this thesis which need mention. Firstly, the use of short follow up periods and the low base rate of sexual recidivism may both have had a significant impact on the findings. The fact that sexual and violent recidivism was used as the outcome for the majority of the studies (due to the low number of sexual recidivists) could have weakened the observed relationship between the psychometrics and recidivism (and also the predictive power of the static tools used within the studies). As such, it would be useful to re-examine the findings using longer follow up periods and thus potentially including a greater number of sexual recidivists. Attempts were made to redress this issue by examining sexual recidivism alone in chapter six. The findings were in fact very similar when using sexual recidivism alone and when using sexual and violent recidivism, but the low number of sexual recidivists means that these findings (from chapter six) are limited themselves. However, it should be noted that the fact that base rates of sexual recidivism are low is positive for community safety, and whether this is a result of interventions provided to these offenders, or changes they are making themselves, this is positive for society.

A further major limitation is the absence of a control group of untreated sexual offenders. This would have enabled comparisons between the predictive accuracy of the psychometric measures in treated and untreated offenders. The difficulties in obtaining a comparable group of untreated sexual offenders, amongst other problems of programme evaluation designs, is well documented (e.g., Hollin, 2008; Marshall & Marshall, 2007). Finding a comparison group is particularly difficult within NOMS as most sexual

offenders are required to complete a treatment programme (if deemed suitable) and may not progress through the Criminal Justice System if they do not do so.

The time of psychometric testing may have also had an impact on the findings. It may be, for example, that improvements observed on the psychometrics pre to post-treatment are predictive of outcome but that these improvements are only predictive for as long as that improvement is sustained. Perhaps a useful methodology for the future would be to examine psychometric scores just prior to recidivism, or at the point of release from prison. It might be that these testing time points may have a closer relationship to recidivism than those taken at the point of treatment, which for some individuals may be a number of years prior to release. This in itself is a further limitation; the fact that individuals had different lengths of time between completing a treatment programme and having psychometric data gathered and their release from prison. This may be a confounding factor and although it was not a focus of the present research, it should be examined in future studies. Furthermore, the sample had all received treatment at different time periods, some as early as 1996, and others in 2006. During this ten year period, the programmes offered to sexual offenders in prison establishments in the UK inevitably underwent a process of maturation and change according to the latest theory and research of the time. As the thesis did not aim to evaluate the effectiveness of the programme in the current research this is not too problematic, but along with changes in the programme there could have been changes in the way the assessments were administered, which could have had an impact on the results. It is worth bearing these issues in mind when drawing conclusions from this thesis.

Further limitations to the design of the empirical chapters of this thesis include a lack of a psychometric measure of sexual interests in some of the studies. This limitation was mainly due to the fact that the main measure of sexual interests currently used, the My Private Interests (MPI) measure, has only been used very recently, and the measure used previously, the Multiphasic Sex Inventory, was only used since 2003 and is copyright, and therefore could not be used for the analysis in chapter six. Future research should attempt to redress this limitation once enough data has been gathered on the MPI. The use of proxy RM2000/s scores for some of the empirical chapters also warrants some discussion. As noted in previous chapters it did not appear that the use of proxy static risk scores had a major impact on the findings (which where possible was tested), but it would have been an improvement to the design if full RM2000/s scores could have been used throughout.

Finally, it is important to mention the fact that for the majority of the studies, a measure of response bias was not used. Although previous research has indicated that socially desirable responding (SDR) may have little impact on psychometric scores for sexual offenders (Mathie & Wakeling, 2011), it would still have been useful to include a measure of response bias in case this affected the predictive validity of the measures. This was not possible in all of the analyses due to the fact that only a smaller subset of the sample had filled in a measure of response bias. As noted in chapter two, there appears to be a complicated relationship between SDR and recidivism, as the findings from this chapter suggested that those scoring higher on a measure of SDR actually were less likely

to be reconvicted of a further offence, and as such SDR may in some offenders be a protective factor. As such, this topic would be worthy of further examination.

Regardless of these limitations, this thesis uses very large samples of sexual offenders to examine the validity of psychometrics for use with this population, and as such makes a significant contribution to the literature on this topic. These samples can also be considered fairly representative of sexual offenders undergoing treatment in the UK Prison system. To the author's knowledge no other study has used such a large sample size to examine the relationship between psychometric variables and recidivism. The thesis also uses a variety of different methodology and statistical analyses to answer the different research questions, which is an advantage of the current research. The contribution of this thesis to the literature therefore does not come from just the findings but also from the different findings using the different methodology and analyses. For example, one conclusion which can be drawn, suggests that within the current samples, contrary to other studies (e.g., Beggs & Grace, 2011), using different methodologies to examine treatment change scores (raw or RCZ scores) does not impact on change scores' predictive validity (see chapter four). The thesis has also contributed to the debate on the use of clinical significant change methodology and the challenge of selecting cut off points within this methodology (see chapter six).

Further Research and Future Directions

Aside from the suggested future research directions already mentioned, which include further examination of the research questions using samples with longer follow ups and

greater recidivism rates, examination of psychometric scores between a treatment and control group, and examination of the impact of time of testing on discriminant and predictive validity, there are other avenues of further research which stem directly from this thesis.

Firstly, it would be useful to examine other potential tools/measures of dynamic risk factors and change within these risk factors using the same sample or a similarly large sample of sexual offenders. Initially it would be useful to examine the predictive validity of the TNA scores gathered as part of SARN by clinicians, both at the pre and post-treatment stage, as well as change in these scores. Psychometric measures contribute to an assessment of the presence of risk factors within the TNA (as described fully in chapter 1), but are only one small part. Thus, although the TNA part of the SARN is not currently used as a predictive tool, it would be interesting to examine the predictive validity of the tool, as it is used in practise. It was not possible to do this in the present thesis as there are currently not big enough samples to do so, but when there are, it would be an extremely worthwhile avenue of research. The present research would suggest that using psychometrics as one part of an assessment of the presence of a risk factor (and not as the only assessment) might be a good strategy, which is how the SARN TNA works. However, the predictive validity of the tool must be examined before firm conclusions about this can be drawn. Evidence from other researchers would indicate that assessments of risk made using a structured professional judgement framework (akin to SARN TNA) are good predictors of outcome (e.g., Olver et al., 2007; Olver & Wong, 2011). However, it should be noted that risk assessment is an inexact science, and

typically when clinicians make judgements about risk, their biases come into play, and in general there may be a tendency towards overestimation of risk (Monahan, 1981). Thus, reliance solely on clinician's ratings of risk may also be problematic.

Equally, it would be useful for future research to examine the use of progress rating scales in measuring treatment change (rather than self-report measures). Previous research is mixed as to the predictive validity of various rating scales. Some research has found rating scales to be unrelated to recidivism (e.g., Barbaree, 2005; Langton, Barbaree, Harkins & Peacock, 2006; Seto & Barbaree, 1999; Quinsey et al., 1998). However, other research has been more promising and has found that such scales can be valid measures of a reduction in risk (e.g., Looman et al., 2005; Marquis et al., 2005; Scalora & Garbin, 2003). Examination of rating scales and other professional judgement scales in routine practise, rather than in a research context, would be worthy of investigation. Beggs and Grace (2011) compared different methodologies for assessing treatment change. Whilst all three methods examined showed some predictive validity, the use of RCZ scores in particular were a favoured method for determining change (psychometrically) and taking into account pre-treatment deviance. However, this methodology may not be easy for clinicians to use in a clinical setting, and these findings were gathered in a research context. Measures of treatment change using ratings by clinicians, such as the VRS:SO and SARN TNA scores may be more easy to apply in a clinical setting.

Second, it would be interesting to gather expert ratings/opinions of the usefulness of psychometrics in the identification and prediction of risk within sexual offenders. Not only would this provide further information on the predictive validity of various other psychometrics, which might have potential, but were not examined in the current thesis, but it might also provide additional information about other ways the measures could be used, how experts believe they should be used within treatment and risk assessment, and how they or their use could be improved. Gathering opinions from a variety of policy makers, researchers, experts and practitioners on topics is a useful way of generating conclusions, and/or consolidating future directions. For example it would be good to examine others' opinions on the use and ability of psychometrics with sexual offenders using some form of the Delphi Method (Linstone & Turoff, 1975) in order to identify and validate agreement between experts and practitioners on this issue. This would provide a useful platform in moving forward with the use of psychometrics in the field of sex offender treatment, in light of the current findings.

Third, it would be useful for future research to focus on unpicking the validity of psychometrics as gathered in different contexts (e.g., used clinically as opposed to being used just within a research context). The present findings provide some indication that the context in which psychometrics are administered and used may have an impact on how valid they are in predicting outcome. For example, when gathered as part of the treatment and risk assessment process (as they are within the NOMS SOTP), individuals know that they are contributing towards their progression through the system, identification of further treatment need and pathways, parole decisions, and even risk

management decisions. Knowing this, individuals may be more likely to respond in a favourable way, particularly at the post-treatment stage. On the other hand, if gathered as part of a research project in which the results do not in any way affect the individuals filling them out, the results may be more likely to have a reliable and valid relationship with reconviction outcome. Examination of the newly devised domain scales (chapter 6) within a research context could provide a useful and interesting avenue for further research and comparison to the present findings.

Fourth, further examination of the issue of change within treatment is needed. It is worth noting that offenders are able to change on dynamic risk factors without attending treatment interventions. In fact, research has shown that some offenders are able to desist from future offending without treatment using their own desistance methods (Maruna, 2001). Thus, measuring change over time without treatment (using a control group) and the relationship between this change and recidivism, would be worthy of investigation. It may be, for example, that other factors aside from these dynamic risk factors are important for some offenders, such as employment, and other social and environmental factors, as well as internal self beliefs. Future research should explore this issue.

Although measuring treatment change is important, it is also critical to look towards theories of desistance and habit change (Maruna, 2001; Prochaska, DiClemente & Norcross, 1992) in making suggestions on future research avenues. From this literature we can conclude that behaviour change is a long term process which evolves in a steady manner. As such, changes on a treatment programme may not necessarily be sustained

beyond imprisonment and on release. Furthermore, treatment gains on a single programme may not be able to account for all the changes within an individual. As such future research should examine change over participation in multiple programmes. There are also other very important issues at play once someone has been released which affect how likely they might be to go on to reoffend. These factors include having available support services, having a job, and having good social support. Thus, even if an individual has made significant changes on a treatment programme, if they do not have sufficient support services available upon release, then these changes might not be observable as a successful outcome in terms of recidivism. Examination of all these factors together is needed before firm conclusions can be made.

Conclusions

In an era where rehabilitation resources are scarce (Scott-Hayward, 2009), there is a danger that treatment developers and providers will be unable to convince policy colleagues on the benefits of providing treatment to sexual offenders if we cannot demonstrate a link between treatment change and reduced risk. It is certainly the case that without robust and valid measures of change that are reliably related to recidivism, evaluations of sexual offender interventions will continue to face difficulty. Having robust and valid measures would enable us to determine who benefits from treatment, how (much) offenders change, when dynamic risk factors have changed enough to have a significant impact on risk, and what impact treatment has. It would be particularly useful to examine what impact treatment has in the wider context. Any measure which can be used to help us answer any of these questions is worthy of investigation. The current

research has contributed to the literature in examining the predictive validity of a large number of psychometric tests in doing just this. Although further research is warranted, the findings indicate that we should not place too much emphasis on psychometric scores and that they should form only one part of an assessment of the presence of dynamic risk factors.

The wider literature and the findings from the current thesis indicate that there is still a lack of evidence regarding measures of change and recidivism. This could be simply that treatment gains are not related to long term recidivism. However, equally, the findings could indicate that the tools used to measure within-treatment progress are not valid. Similarly, it appears that the use of different methodology to assess treatment change can have a significant impact on the results. It is thus extremely important to establish the reliability and validity (particularly predictive validity) of the tools we use to measure dynamic risk factors and to measure treatment change before they are routinely implemented in a clinical setting.

REFERENCES

Abel, G. G., Becker, J. V., & Cunningham-Rathner, J. (1984). Complications, consent, and cognitions in sex between children and adults. *International Journal of Law and Psychiatry, 7,* 89-103.

Abel, G. G., Gore, D. K., Holland, C. L., Camp, N., Becker, J. V., & Rathner, J. (1989). The measurement of the cognitive distortions in child molesters. *Annals of Sex Research, 2,* 135-152.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723. doi: 10.1109/TAC.1974.1100705

Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 347-367. doi: 10.1177/107906320701900402

Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory–Revised*. Toronto, ON, Canada: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: LexisNexis.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: LexisNexis, Matthew Bender.

Andrews, D. A., Kiessling, J. J., Mickus, S. G., & Robinson, R. (1986). *The Level of Supervision Inventory: Second Report.* Toronto: Ministry of Correctional Service of Ontario.

Bakker, L., Hudson, S., Wales, D., & Riley, D. (1998). *And there was light: Evaluating the Kia Marama treatment programme for New Zealand sex offenders against children*. Psychological Service Report, Department of Corrections. Wellington, New Zealand.

Barbaree, H. E. (2005). Psychopathy, treatment behavior, and recidivism: an extended follow-up of Seto and Barbaree. *Journal of Interpersonal Violence, 20*, 1115-31.

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*, 490-521.

Barnett, G. D., & Mann, R. E. (2011). Good Lives and Risk Assessment: Collaborative Approaches to Risk Assessment with Sexual Offenders. In: Kemshall, H. and Wilkinson, B. (eds.) *Good Practice in Assessing Risk: Current Knowledge, Issues and Approaches*. London: JKP.

Barnett, G. D., Wakeling, H. C., & Howard, P. D. (2010). An examination of the predictive validity of the Risk Matrix 2000 in England and Wales. *Sexual Abuse: A Journal of Research and Treatment*, *22*, 443-470. doi: 10.1177/1079063210384274

Barnett, G. D., Wakeling, H. C., Mandeville-Norden, R., & Rakestrow, J. (2012). How useful are psychometric scores in predicting recidivism for treated sex offenders? *International Journal of Offender Therapy and Comparative Criminology*, *56*, 420-446.

Barnett, G. D., Wakeling, H. C., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime and Law*, *19*, 85-110.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The Analysis and Interpretation of Multivariate Data for Social Scientists*. Boca Raton, Florida: Chapman & Hall.

Beckett, R. C. (1987). *Children and Sex Questionnaire.* Oxford, UK: Oxford Forensic Services.

Beckett, R. C., Fisher, D., Mann, R. E., & Thornton, R. E. (1997). The Relapse Prevention Questionnaire and Interview. In H. Eldridge (Ed.), *Therapists Guide for Maintaining Change: Relapse Prevention Manual for Adult Male Perpetrators of Child Sexual Abuse.* California, US: Sage Publications.

Beech, A. R. (1998). A psychometric typology of child abusers. *International Journal of Offender Therapy and Comparative Criminology*, *42*, 319-339. doi: 10.1177/0306624X9804200405

Beech, A. R., Bishopp, D., & Fisher, D. (2004, October). *Pre-post change designs*. Paper presented at the 23rd Annual Conference of the Association for the Treatment of Sexual Abusers, Alberquerque, New Mexico, USA.

Beech, A. R., Erikson, M., Friendship, C., & Ditchfield, J. (2001). A six-year follow-up of men going through probation-based sex offender treatment programmes. *Research Findings, 144*. Home Office Research, Development and Statistics Directorate.

Beech, A. R., & Fisher, D. (1997). Assessment of clinically significant change in treated child abusers. *Proceedings of the East-West Conference: Child Sexual Abuse and Sexual Violence*. Prague, Czech Republic, September 1996.

Beech, A. R., Fisher, D., & Beckett, R. (1999). *STEP3: An evaluation of the prison sex offender treatment programme.* Home Office Occasional Paper. London: Home Office.

Beech, A. R., Fisher, D. D., Thornton, D. (2003). Risk assessment of sex offenders. *Professional Psychology: Research and Practice, 34*, 339-352.

Beech, A. R., & Ford, H. (2006). The relationship between risk, deviance, treatment outcome and sexual reconviction in a sample of child sexual abusers completing residential treatment for their offending. *Psychology, Crime and Law*, *12*, 685-701. doi: 10.1080/10683160600558493

Beech, A. R., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of UK child abusers. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 155-167. doi: 10.1023/A:1014672231744

Beech, A. R., & Hamilton-Giachritsis, C. E. (2005). Relationship between therapeutic climate and treatment outcome in group-based sexual offender treatment programs. *Sexual Abuse: A Journal of Research and Treatment, 17*, 127-140.

Beech, A. R., & Ward, T. (2004). The integration of etiology and risk in sexual offenders: A theoretical framework. *Aggression and Violent Behavior, 10*, 31-63.

Beggs, S. (2010). Within-treatment outcome among sexual offenders: A review. *Aggression and Violent Behavior, 15, 369-379.*

Beggs, S., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity. *Journal of Consulting and Clinical Psychology, 79, 182-192.*

Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles, CA: BMDP Statistical Software.

Bickley, J. A., & Beech, R. (2003). Implications for treatment of sexual offenders of the Ward and Hudson model of relapse. *Sexual Abuse: A Journal of Research and Treatment, 15,* 121-134.

Blackburn, R. (1982). *The Special Hospital Assessment of Personality and Socialization (SHAPS)*. Unpublished manuscript.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk – 20: Professional guidelines for assessing risk of sexual violence.* Vancouver, B.C.: The Mental Health, Law, and Policy Institute.

Bowen, E., Gilchrist, E. A., & Beech, A. R. (2005). An examination of the impact of the community-based rehabilitation on the offending behaviour of male domestic violence offenders and the characteristics associated with recidivism. *Legal and Criminological Psychology*, *10*, 189-209. doi: 10.1348/135532505X36778

British Psychological Society (2009). Code of Ethics and Conduct. Guidance published by the Ethics Committee of the British Psychological Society. ISBN: 978-1-85433-495-4.

Bumby, K.M. (1996). Assessing the Cognitive Distortion of Child Molesters and Rapists: Development and Validation of the MOLEST and RAPE Scales. *Sexual Abuse: A Journal of Research and Treatment, 8,* 37-54.

Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology, 38*, 217-230.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology*, *21*, 343-349.

Buss, A. H., & Perry, M. P. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459.

Cann, J., Friendship, C., & Gozna, L. Assessing crossover in a sample of sexual offenders with multiple victims. *Legal and Criminological Psychology, 12,* 149-163.

Caprara, G. V. (1986). Indications of aggression: The dissipation-rumination scale. *Personality and Individual Differences*, *7*, 763-769. doi: 10.1016/0191-8869(86)90074-7

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, I*, 245-276.

Christensen, L., & Mendoza, J.L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 15*, 305-308.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1992). Quantitative Methods in Psychology. A Power Primer. *Psychological Bulletin, 112,* 155-159.

Collaborative Outcome Data Committee. (2007). Guidelines for the Evaluation of Sexual Offender Treatment Outcome Research (CODC Guidelines), *User Report 2007-02. Ottawa: Public Safety and Emergency Preparedness Canada.* Available from: www.publicsafety.gc.ca/res/cor/rep/cprmindex-eng.aspx.

Cooke, D. J., & Mitchie, C. (2014). The generalizability of the Risk Matrix 2000: On model shrinkage and the misinterpretation of the area under the curve. *Journal of Threat Assessment and Management, 1(1)*, 42-55.

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research and Evaluation, 10*, ISSN 1531-7714.

Cortina, J. M. (1993). What is Coefficient Alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104.

Cortoni, F., & Marshall, W. L. (2001). Sex as a coping strategy and its relationship to juvenile sexual history and intimacy in sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *13*, 27-43. doi: 10.1023/A:1009562312658

Craig, L. A., Beech, A. R., & Browne, K. D. (2006). Cross-validation of the Risk Matrix 2000 sexual and violent scales. *Journal of Interpersonal Violence*, *21*, 612-633. doi: 10.1177/0886260506286876

Craig, L. A., Thornton, D., Beech, A. R., & Browne, K. D. (2007). The relationship of statistical and psychological risk markers to sexual reconviction in child molesters. *Criminal Justice and Behavior*, *34*, 314-329. doi: 10.1177/0093854806291416

Craissati, J., & Beech, A. (2003). A review of dynamic variables and their relationship to risk prediction in sex offenders. *Journal of Sexual Aggression, 9,* 41–55.

Davis, M. H. (1980). A multi-dimensional approach to individual differences in empathy. *JSAS Catalogue of Selected Documents in Psychology*, *10*, 85-100.

Dempster, R. J., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and non recidivists. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 121-138. doi: 10.1023/A:1014668130835

Douglas, K. S., Epstein, M. E., & Poythress, N. G. (2008). Criminal recidivism among juvenile offenders: testing the incremental and predictive validity of three measures of psychopathic features. *Law and Human Behavior, 32*, 423-438.

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy and Law*, 11, 347-383.

Dutton, D. G., Saunders, K., Starzomski, A., & Bartholomew, K. (1994). Intimacy-anger and insecure attachment as precursors of abuse in intimate relationships. *Journal of Applied Social Psychology, 24,* 1367-87.

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). *Social Problem-Solving Inventory – Revised.* Technical Manual. MHS.

Edens, J. F., Hart, S. D., Johnson, D. W., Johnson, J. K., & Olver, M. E. (2000). Use of the Personality Assessment Inventory to asses psychopathy in offender populations. *Psychological Assessment*, *12*, 132-139.

Eysenck, S. B. G., & Eysenck, H. J. (1978). Impulsivity and venturesomeness: Their place in a dimensional system of personality description. *Psychological Reports*, *43*, 1247-1255.

Farren, L., & Barnett, G. (2014). Measuring offence-related sexual interests in convicted sexual offenders. An initial test of the reliability and validity of a self-report scale. NOMS Analytical Summary. ISBN 978 [1-84099-650-0]

Fanniff, A. M., & Becker, J. V. (2006). Specialized assessment and treatment of adolescent sex offenders. *Aggression and Violent Behavior, 11,* 265-282.

Finkelhor, D. (1984). *Child Sexual Abuse: New Theory and Research* (http://www.amazon.com/exec/obidos/ISBN=0029100208/instituteforpsychttp://www.amazon.com/exec/obidos/ISBN=0029100208/instituteforpsyc). New York: Free Press.

Floyd, F. J., & Widaman, K. F. (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment, 7*, 286-299.

Friendship, C., Falshaw, L., & Beech, A.R. (2003). Measuring the real impact of accredited offending behavior programmes. *Legal and Criminological Psychology*, *8*, 115-127.

Friendship, C., Mann, R. E., & Beech, A. R. (2003). Evaluation of a National Prison-based Treatment Program for Sexual Offenders in England and Wales. *Journal of Interpersonal Violence*, *18*, 744-759. doi: 10.1177/0886260503253236

Frude, N., Honess, T., & Maguire, M. (1994). *CRIME-PICS II Manual.* Cardiff, Wales: Michael and Associates.

Furby, L., Weinrott, M. R., & Blackshaw, L. (1989). Sex offender recidivism: a review. [Review]. *Psychological Bulletin, 105,* 3-30.

Grady, M. D., Broderson, M., & Abramson, J. M. (2011). The state of psychological measures for adult sexual offenders. *Aggression and Violent Behavior, 16*, 227-240.

Grubin, D. (2008). *Validation of Risk Matrix 2000 for use in Scotland.* Report prepared for the Risk Management Authority. Retrieved 21 January 2011 from http://rmascotland.org/ViewFile.aspx?id=329

Hanson, R. K. (1996). How to know what works with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 9,* 129-145.

Hanson, R. K. (1998).What do we know about sexual offender risk assessment. *Psychology, Public Policy, and Law*, *4*, 50–72.

Hanson, K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The Principles of Effective Correctional Treatment Also Apply to Sexual Offenders: A Meta-analysis. *Criminal Justice and Behavior, 36*, 865-891.

Hanson, R.K., & Bussière, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66,* 348-362.

Hanson, R. K., Cox, B. J., & Woszczyn, C. (1991). Assessing treatment outcome for sexual offenders. *Annals of Sex Research*, *4*, 177-208. doi: 10.1007/BF00850052

Hanson, R. K., Gizzarelli, R., & Scott, H. (1994). The attitudes of incest offenders: Sexual entitlement and acceptance of sex with children. *Criminal Justice and Behavior*, *27*, 6-35. doi: 10.1177/BF00850052

Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*¸ *13*, 105-122. doi: 10.1023/A:1026600304489

Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism amongst sexual offenders: Amulti-site study of Static-2002. *Law and Human Behavior, 34,* 198–211. doi:10.1007/s10979-009-9180-1

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, *73*, 1154-1163. doi: 10.1037/0022-006X.73.6.1154

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*, 1-21. doi: 10.1037/a0014421

Hanson, R, K., Steffy, R. A., & Gauthier, R. (1993). Long-term recidivism of child molesters. *Journal of Consulting and Clinical Psychology*, *61*, 646-652. doi: 10.1037//0022-006X.61.4.646.

Hanson, R.K. & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119-136. doi: 10.1023/A:1005482921333

Hanson, R. K., & Wallace-Capretta, S. (2000). *Predicting recidivism among male batterers.* Toronto: Public Works and Government Services Canada.

Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist–Revised*. Toronto, ON, Canada: Multi-Health Systems.

Harkins, L (2008). *Sex offender treatment: Risk, need and responsivity and process issues*. Unpublished PhD thesis, University of Birmingham, UK.

Harkins, L., & Beech, A. (2007). Measurement of the effectiveness of sex offender treatment. *Aggression and Violent Behavior, 12*, 36-44.

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in Developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, *15*, 361-387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Harris, G. T. & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. In R. A. Prentky, E.S. Janus, & M. C. Seto (Eds.) *Understanding and managing sexually coercive behavior.* (pp 198-210). New York: Annals of the New York Academy of Sciences, Vol. 989.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, *20*, 315-335.

Hart, S. D., Kropp, P. R. & Laws, D. R. (2003). *Risk for Sexual Violence Protocol (RSVP): Structured Professional Guidelines for Assessing Risk of Sexual Violence.* Burnaby, BC: The Mental Health, Law and Policy Institute, Simon Fraser University.

Hedderman, C., & Sugg, D. (1996). Does treating sex offenders reduce reoffending? *Research Findings*, 45. Home Office Research, Development and Statistics Directorate. Available online at: www.homeoffice.gov.uk/rds/pdfs/r45.pdf

Helmus, L., Hanson, R. K., Babchishin, K. M., & Mann, R. E. (2013). Attitudes supportive of sexual offending predict recidivism: A meta-analysis. *Trauma, Violence & Abuse, 14,* 34-53.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement, 66*, 393-416.

Hogue, T. E. (1994). Goal attainment scaling: A measure of clinical impact and risk assessment. *Issues in Criminological and Legal Psychology, 21,* 96-102.

Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology, 63,* 272-279.

Hollin, C. (2008). Evaluating offending behavior programs: Does only randomization glister? *Criminology and Criminal Justice, 8,* 89-106.

Hood, R., Shute, S., Feilzer, M., & Wilcox, A. (2002). Sex offenders emerging from long-term imprisonment. *British Journal of Criminology*, 42, 371-394. doi: 10.1093/bjc/42.2.371

Howard, P., Francis, B., Soothill, K., & Humphreys, L. (2009). OGRS 3: the revised offender group reconviction scale. *Ministry of Justice Research Summary, 7/09*, available at: www.justice.gov.uk/publications/docs/oasys-research-summary-07-09-ii.pdf

Hudson, S. M., Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic risk factors: The Kia Marama Evaluation. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 103-119. doi: 10.1023/A:1014616113997

Jacobsen, N.S., Roberts, L.J., Berns, S.B., & McGlinchey, J.B. (1999). Methods for determining the clinical significance of treatment effects: Description, application and alternatives. *Journal of Consulting and Clinical Psychology*, *67*, 300-307.

Jacobsen, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.

Jöreskog, K. G., & Sörbom, D. (1989). *Lisrel 7. A guide to the program and applications*. (2nd Ed.). Chicago, Illinois: SPSS Inc.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kazdin, A. E. (2003). Clinical significance: Measuring whether interventions make a difference. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research, third edition* (pp. 691 -710). Washington, DC: American Psychological Association.

Keeling, J. A., Rose, J. L., & Beech, A. R. (2006). An investigation into the effectiveness of a custody-based cognitive-behavioral treatment for special needs sexual offenders. *Journal of Forensic Psychiatry and Psychology*, *17*, 372-392.

Kline, P. (1993). *The Handbook of Psychological Testing*. London, Routledge.

Kline, P. (2000). *The Handbook of Psychological Testing.* Routledge.

Knight, R. A., & Thornton, D. (2007). Evaluating and improving risk

assessment schemes for sexual recidivism: A long-term followup of convicted sexual offenders (Document No. 217618). U.S. Department of Justice.

Kroner, D., & Loza, W. (2001). Evidence for the efficacy of self-report in predicting violent and non-violent criminal recidivism. *Journal of Interpersonal Violence*, *16*, 168-177.

Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior*, *28*, 471-489.

Kroner, D. G., & Weekes, J. R. (1996). Balanced inventory of desirable responding: Factor structure, reliability, and validity with an offender sample. *Personality and Individual Differences, 21,* 323-333.

Langton, C. M., Barbaree, H. E., Harkins, L., & Peacock, E. J. (2006). Sex offenders' response to treatment and its association with recidivism as a function of psychopathy. *Sexual Abuse: A Journal of Research and Treatment, 18,* 99-120.

Langton, C. M., Barbaree, H. E., Hansen, K. T., Harkins, L., & Peacock, E. J. (2007). Reliability and validity of the Static-2002 among adult sexual offenders with reference to treatment status. *Criminal Justice and Behavior, 34*, 616-640.

Levenson, H. (1974). Multidimensional locus of control in prison inmates. *Personality and Social Psychology Bulletin*, *1*, 354-356. doi: 10.1111/j.1559-1816.1975.tb00686.x

Linstone, H. A., & Turoff, M. T. (1975). *The Delphi Method: techniques and applications.* London, Addison-Wesley.

Looman, J., Abracen, J., Serin, R., & Marquis, P. (2005). Psychopathy, treatment change, and recidivism in high-risk, high-need sexual offenders. *Journal of Interpersonal Violence*, *20*, 549-568.

Loza, W. (1996). Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and nonviolent recidivism. *Unpublished manuscript.*

Loza, W., Cumbleton, A., Shahinfar, A., Neo, L. H., Evans, M., Conley, M., & Summers, R. (2004). Cross-validation of the Self-Appraisal Questionnaire (SAQ): An offender risk and need assessment measure on Australian, British, Canadian, Singaporean, and American Offenders. *Journal of Interpersonal Violence*, *19*, 1172-1190.

Loza, W., & Loza-Fanous, A. (2000). Predictive validity of the Self-Appraisal Questionnaire (SAQ): A measure for assessing violent and nonviolent release failures. *Journal of Interpersonal Violence*, *15*, 1183-1191.

Loza, W., & Loza-Fanous, A. (2003). More evidence for the validity of the Self-Appraisal Questionnaire (SAQ) for predicting violent and nonviolent recidivism: A five year follow up study. *Criminal, Justice and Behavior*, *30*, 709-721.

Lusignaan, R., & Marleau, J. D. (2007). Risk assessment and offender victim relationship in juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology, 51,* 433-443.

Mak, A. S. (1991). Psychosocial control characteristics of delinquents and nondelinquents. *Criminal Justice and Behavior, 18,* 297-303. doi: 10.1177/0093854891018003003

Mandeville-Norden, R., Beech, A., & Hayes, E. (2008). Examining the effectiveness of a UK community-based sexual offender treatment programme for child abusers. *Psychology, Crime and Law, 14,* 493-512.

Mann, R. E., & Fernandez, Y. M. (2006). Sex offender programmes: Concept, theory, and practice. In C.R. Hollin, & E.J. Palmer (Eds.), *Offending behaviour programmes: Development, application, and controversies* (pp. 155-177). New York: Wiley.

Mann, R. E., Hanson, R. K. & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment, 22*, 172-190. doi: 10.1177/1079063210366039

Mann, R. E., & Thornton, D. (1998). The evolution of a multi-site sex offender treatment programme. In W.L. Marshall, Y.M. Fernandez, S.H. Hudson, & T. Ward (Eds.), *Sourcebook of treatment programs for sexual offenders* (pp. 47-58). New York: Plenum Press.

Mann, R. E., Webster, S., Wakeling, H., & Marshall, W. (2007). The measurement and influence of child sexual abuse supportive beliefs. *Psychology, Crime and Law*, *13*, 443-458. doi: 10.1080/10683160601061141

Marques, J.K., Wideranders, M., Day, D.M., Nelson, C., & van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Program (SOTEP). *Sexual Abuse: A Journal of Research and Treatment*, *17*, 79-107.

Marshall, W. L. (1993). The role of attachment, intimacy and loneliness in the etiology and maintenance of sexual offending. *Sexual and Marital Therapy, 8,* 109-121.

Marshall, W. L. (1995). *Sex with Children is Justifiable Questionnaire.* Unpublished.

Marshall, W. L., & Marshall, L. E. (2007). The utility of the Random Controlled Trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy? *Sexual Abuse: A Journal of Research and Treatment, 19*, 175-191.

Marshall, W. L., & Marshall, L. E. (2010). Attachment and intimacy in sexual offenders: An update. *Sexual and Relationship Therapy*, *25*, 86-90.

Marshall, W. L., Serran, G. A., Fernandez, Y. M., Mulloy, R., Mann, R. E., & Thornton, D. (2003). Therapist characteristics in the treatment of sexual offenders: Tentative data on their relationship with indices of behaviour change. *Journal of Sexual Aggression, 9*, 25-30.

Marshall, W. L., Serran, G. A., Moulden, H., Mulloy, R., Fernandez, Y. M., Mann, R. E., & Thornton, D. (2002). Therapist features in sexual offender treatment: Their reliable identification and influence on behaviour change. *Clinical Psychological and Psychotherapy, 9,* 395-405.

Maruna, S. (2001). *Making good: How Ex-Convicts Reform and Rebuild their Lives.* Washington, American Psychological Association.

Marziano, V., Ward, T., Beech, A., & Pattison, P. (2006). Identification of five fundamental implicit theories underlying cognitive distortions in child abusers: A preliminary study. *Psychology, Crime and Law*, *12*, 97-105. doi: 10.1080/10683160500056887

Mathie, N., & Wakeling, H. C. (2011). Assessing socially desirable responding and its impact on self-report measures among sexual offenders. *Psychology, Crime and Law*, *17*, 215-238.

McDougall, C., Petty, A. E., Clarbour, J., Bowles, R., & Worthy, G. (2009). Evaluation of HM Prison Service Enhanced Thinking Skills Programme. Report on the outcomes from a randomized controlled trial. *Ministry of Justice Research Series, 3/09.* Retrieved 21 January 2011 from http://www.justice.gov.uk/publications/docs/report-on-the-outcomes-from-a-randomised-controlled-trial1.pdf.

McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2012). The Sex Offender Treatment Intervention and Progress Scale (SOTIPS) Psychometric Properties and Incremental Predictive Validity with Static-99R. *Sexual Abuse: A Journal of Research and Treatment, 24,* 431-458.

Michael, P. G., & Li, S. T. (2010). Psychometrics. In *Handbook of Clinical Psychology Competencies*, pp. 343-365. Springer New York.

Miller, R. S., & Lefcourt, H. M. (1982). The assessment of social intimacy. *Journal of Personality Assessment, 46*(5), 514-518.

Mills, J. F., & Kroner, D. G. (2006). Impression management and self-report among violent offenders. *Journal of Interpersonal Violence, 21*, 178-192.

Mills, J. F., Kroner, D. G., & Hemmati, T. (2004). The Measures of Criminal Attitudes and Associates (MCAA): The prediction of general and violent recidivism. *Criminal Justice and Behavior*, *31*, 717-733.

Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence for the robustness of self-report among offenders. *Criminal Behaviour and Mental Health, 13*, 140-150.

Monahan, J. (1981). *The clinical prediction of violent behavior*. Washington, DC: Government Printing House.

Mosher, D. L., & Sirken, M. (1984). Measuring a Macho Personality Constellation. *Journal of Research in Personality, 18*, 150-163.

Motuik, M. S., Motuik, L. L., & Bonta, J. (1992). A comparison of self-report and interview-based inventories in offender classification. *Criminal Justice and Behavior*, *19*, 143-159.

Nichols, H. R., & Molinder, I. (1984). *Multiphasic Sex Inventory.* User Manual. Nichols and Molinder Assessments.

NOMS Interventions Services Group (unpublished). Women are deceitful scale. Available from Interventions Services, NOMS, 4th Floor Clive House, 70 Petty France, London, UK.

Nunes, K. L. Babchishin, K. M., & Cortoni, F. (2011). Measuring treatment change in sex offenders: Clinical and statistical significance. *Criminal Justice and Behavior, 38*, 157-173. doi: 10.1177/0093854810391054

Nunes, K. L., Hermann, C. A., Maimmone, S., & Woods, M. (2014). Thinking Clearly About Violent Cognitions: Attitudes May Be Distinct From Other Cognitions. *Journal of Interpersonal Violence.* Published online 4 July 2014 DOI: 10.1177/0886260514540329

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory.* (3rd Ed.) New York: McGraw-Hill.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers, 32,* 396-402.

Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. P. (2013). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale-Sexual Offender version. *Sexual Abuse: A Journal of Research and Treatment.* Published online DOI: 10.1177/1079063213502679

Olver, M. E., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. C. P. (in press). A Psychometric Examination of Treatment Change in a Multisite Sample of Treated Canadian Federal Sexual Offenders. *Submitted to Law and Human Behavior.*

Olver, M. E., Nicholaichuk, T. P., & Wong, S. C. P. (24 July 2013a). The predictive and convergent validity of a psychometric battery used to assess sex offenders in a treatment program: An 18-year follow-up. *Journal of Sexual Aggression.* Published online DOI: 10.1080/13552600.2013.816791

Olver, M. E., & Wong, S. C. P. (2011). A Comparison of Static and Dynamic Assessment of Sexual Offender Risk and Need in a Treatment Context. *Criminal Justice and Behavior 38*, 113-126**.**

Olver, M. E., Wong, S. C. P., Nicholaichuk, T. P., & Gordon, A. E. (2007). The validity and reliability of the Violence Risk Scale – Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318-329.

Olver, M. E., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. P. (in press). A psychometric examination of treatment change in a multisite sample of treated Canadian federal sexual offenders. *Law and Human Behavior.*

Ones, D.S., Viswesvaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81(6),* 660-679. doi : 1037/0021-9010.81.6.660.

Otter, Z., & Egan, V. (2007). The evolutionary role of self-deceptive enhancement as a protective factor against antisocial cognition. *Personality and Individual Differences, 43(8),* 2258-2269.

Paulhus, D. L. (1984). Two Component Models of Socially Desirable Responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74,* 1197–1208.

Prochaska, J., DiClemente, C., & Norcross, J. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 1002-1114.

Proulx, J., Pellerin, B., Paradis, Y., McKibben, A., Aubut, J., & Ouimet, M. (1997). Static and dynamic predictors of recidivism in sexual aggressors. *Sexual Abuse: A Journal of Research and Treatment*, 9, 7-27.

Quinsey, V. L., Khanna, A., & Malcolm, P. B. (1998). A retrospective evaluation of the regional treatment centre sex offender treatment program. *Journal of Interpersonal Violence*, *13*, 621-644.

Rallings, M., & Webster, S.D. (2001). The psychometric properties of the HMPS SOTP psychometric battery. Unpublished raw data. Available from Interventions Services, NOMS, 4th Floor Clive House, 70 Petty France, London, UK.

Raynor, P. (1998). Attitudes, Social problems and reconvictions in the 'STOP' Probation Experiment. *Howard Journal of Criminal Justice, 37*, 1-15.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC, Cohen's *d* and *r*. *Law and Human Behavior, 29*, 615-620.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84(5)*, 754-775. doi:10.1037/0021-9010.84.5.754.

Roger, D., & Najarian, B. (1989). The Construction and Validation of a New Scale for Measuring Emotion Control. *Personality and individual Differences, 10,* 845-853.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.), New York: McGraw-Hill.

Ross, R. R., & Fabiano, E. A. (1985). Time to Think: A Cognitive Model of Delinquency Prevention and Rehabilitation. Johnson City, TN: Institute of Social

Services and Arts Inc. (available from R. R. Ross, Department of Criminology, University of Ottawa).

Russell, D., Peplan, L. A., & Cutrona, C. A. (1980). The revised UCLA loneliness scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*, *39*, 472-480. doi:10.1037/0022-3514.39.3.472

Rust, J. & Golombok, S. (1999). *Modern Psychometrics: The science of psychological assessment*, 2nd edition, Routledge, London & New York.

Scalora, M. J., & Garbin, C. (2003). A multivariate analysis of sex offender recidivism. *International Journal of Offender Therapy and Comparative Criminology, 47, 309-323.*

Schretlan, D., & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation. *Behavioral Sciences and the Law*, *8*, 75-84.

Scott-Hayward, C. S. (2009). *The Fiscal Crisis in Corrections Rethinking Policies and Practices.* Vera, Institute of Justice.

Serin, R. C. & Kennedy, S. (1997). *Treatment readiness and responsivity: Contributing to effective correctional programming.* (Research Report. R-54). Ottawa: Correctional Services of Canada.

Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and Violent Behavior, 18,* 32-53.

Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC: American Psychological Association.

Seto, M. C., & Barbaree, H. E. (1999). Psychopathy, treatment behavior and sex offender recidivism. *Journal of Interpersonal Violence*, *14*, 1235-1248.

Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology*, *63*, 1044-1048.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., & Sydeman, S. J. (1994). State-Trait Anger Inventory and State-Trait Anger Expression Inventory. In *The use of psychological testing for treatment planning and outcomes assessment* (ed. M. E. Maurish.), pp. 300–321. Lawrence Erlbaum Associates: New Jersey.

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 139-154. doi: 10.1023/A:1014620214905

Thornton, D. (2007). *Scoring Guide for the Risk Matrix 2000.9/SVC: February 2007 Version.* Retrieved 18 October 2010 from http://www.cfcp.bham.ac.uk/Extras/SCORING%20GUIDE%20FOR%20RISK%20MATRIX%202000.9-%20SVC%20-%20(ver.%20Feb%202007).pdf

Thornton, D. (2010). *Scoring Guide for the Risk Matrix 2000.10/SVC.*

Thornton, D., Beech, A., & Marshall, W. L. (2004). Pretreatment Self-Esteem and Posttreatment Sexual Recidivism. *International Journal of Offender Therapy and Comparative Criminology*, *48*, 587-599.

Thornton, D., & Knight, R. A. (2013). Construction and Validation of SRA-FV Need Assessment. *Sexual Abuse: A Journal of Research and Treatment.* Available online: Published online before print December 30, 2013, doi: 10.1177/1079063213511120

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing between and combining risks for sexual and violent recidivism. In R. Prentky, E. Janus, M. Seto, & A. W. Burgess (Eds), *Understanding and managing sexually coercive behavior, Annals of New York Academy of Sciences*, *989*, 225-235. doi: 10.1111/j.1749-6632.2003.tb07308.x

Thornton, D. & Travers, R. (1991). Longitudinal Study of the Criminal Behaviour of Convicted Sex Offenders. In*: Proceedings of the Prison Psychologists Conference*. London: H.M. Prison Service.

Tierney, D. W., & McCabe, M. P. (2001). An evaluation of self-report measures of cognitive distortions and empathy among Australian sex offenders. *Archives of Sexual Behavior, 30,* 495-519.

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*, 287-316.

Underhill, J., Wakeling, H. C., Mann, R. E., & Webster, S. (2008). Male sexual offenders' emotional openness with men and women. *Criminal Justice and Behavior*, *35*, 1156-1173. doi: 10.1177/0093854808320268

Wakeling, H. C. (2007). The Psychometric Validation of the Social Problem-Solving Inventory-Revised with UK Incarcerated Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment, 19*, 217-236.

Wakeling, H. C., & Barnett, G. D. (2011). Measuring grievance thinking in sexual offenders: The revised dissipation-rumination scale. *Journal of Sexual Aggression,* 17, 273-289.

Wakeling, H. C., & Barnett, G. D. (2014). The relationship between psychometric test scores and reconviction in sexual offenders undertaking treatment. *Aggression and Violent Behavior, 19,* 138-145.

Wakeling, H. C., Beech, A. R., & Freemantle, N. (2013). Investigating treatment change and its relationship to recidivism in a sample of 3773 sex offenders in the UK. *Psychology, Crime and Law, 19*, 233-252.

Wakeling, H. C., Freemantle, N., Beech, A. R. & Elliott, I. (2011). Identifying predictors of recidivism in a large sample of UK sexual offenders: A prognostic model. *Psychological Services, 8,* 307-318.

Wakeling, H. C., & Travers, R. (2010). Evaluating offending behaviour programmes in prison. In J.M. Brown and E.A. Campbell (Eds.) *The Cambridge Handbook of Forensic Psychology*, pp. 820-829. Cambridge University Press.

Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes. *Criminal Justice and Behavior*, *33*, 179-304.

Walters, G. D., White, T. W., & Denney, D. (1991). The Lifestyle Criminality Screening Form: Preliminary data. *Criminal Justice and Behavior*, *18*, 406-418.

Wampold, B.E. (2001). *The great psychotherapy debate: Models, methods and findings.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ward, T. (2000). Sexual offenders' cognitive distortions as implicit theories. *Aggression and Violent Behaviour: A Review Journal*, *5*, 491-507. doi: 10.1016/S1359-1789(98)00036-6

Ward, T. (2002) 'Good lives and the rehabilitation of offenders: promises and problems.' *Aggression and Violent Behavior 7*, 513–528.

Ward, T., & Brown, M. (2004). The Good Lives Model and conceptual issues in offender rehabilitation. *Psychology, Crime and Law, 10*, 243–257.

Ward, T., & Keenan, T. (1999). Child molesters' implicit theories. *Journal of Interpersonal Violence*, *14*, 821-838. doi: 10.1177/088626099014008003

Ward, T., & Marshall, W. L. (2004). Good lives, aetiology and the rehabilitation of sex offenders: a bridging theory. *Journal of Sexual Aggression, 10*, 2, 153–169.

Webster, C. D., Eaves, D., Douglas, K. S., & Wintrup, A. (1995). *The HCR-2 Scheme: The assessment of dangerousness and risk.* Vancouver, Canada: Simon Fraser University and British Columbia Forensic Psychiatric Services.

Webster, S., Mann, R. E., Thornton, D., & Wakeling, H. C. (2007). Further validation of the Short Self-esteem Scale with sexual offenders. *Legal and Criminological Psychology*, *12*, 207-216. doi: 10.1348/135532506X114347

Webster, S. D., Mann, R. E., Carter, A., Long, J., Milner, R., O'Brien, M. D., Wakeling, H. C., & Ray, N. L. (2006). Inter-rater reliability of dynamic risk assessment in sexual offenders. *Psychology, Crime and Law, 12*, 439-452.

Wilkinson, J. (2005). Evaluating evidence for the effectiveness of the Reasoning and Rehabilitation Programme. *The Howard Journal*, *44*, 70-85.

Williams, F. (2007). *My Private Interests Measure for sexual offenders with cognitive and/or social deficits.* Unpublished manuscript.

Wilson, G. (1978). *The secrets of sexual fantasy*. London: J. M. Dent & Sons.

Wise, E.A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, *82*, 50-59.

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the violence risk scale: A treatment friendly violence risk assessment scale. Psychology, Public Policy and Law. 12, 279-309. doi:10.1037/1076-8971.12.3.279

Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). The Violence Risk Scale: Sexual Offender version (VRS-SO). Regional Psychiatric Centre and University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Zwick, W. R., & Velicer, W. F. (1986). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

Appendices

Appendix A: Ethical Approval

Appendix B: Description of Psychometrics

Old Battery Measures

The Entitlement to sex scale (Hanson, Gizzarelli & Scott, 1994) measures whether an individual believes they are entitled to sex. The scale has 9 items scored on a 5 point likert scale, with higher scores representing a greater endorsement of entitlement attitudes. The scale has adequate reliability; the internal consistency of the scale is α = .65, and the test-retest reliability of the scale is r = .67 (Rallings & Webster, 2001).

The Impulsivity scale (Eysenck & Eysenck, 1978) has 13 items and examines a tendency to act impulsively, without regard for long-term consequences. Item responses are 'yes' or 'no', and scale scores range from 0 to 26. The scale has adequate reliability: internal consistency (α = .84) and test-retest reliability (.79; Rallings & Webster, 2001).

The Interpersonal Reactivity Index (IRI: Davis, 1980) has 28 items measuring the cognitive and emotional components of empathy. The response format is a 5 point likert scale ranging from 0 (does not describe me well) to 4 (describes me very well). There are four subscales each with 7 items, and scores ranging from 0 to 28. The four subscales have adequate reliability (Rallings & Webster, 2001): internal consistency (fantasy, (α = .77), empathic concern (α = .72), perspective taking (α = .72), and personal distress (α = .74)); and test re-test reliability (fantasy, r=.77; empathic concern, r=.79; perspective taking, r=.81; and personal distress, r=.74).

The Locus of Control (Levenson, 1974) questionnaire examines the extent to which an individual equates his experiences to internal or external factors. It has 18 items, and is scored on a five point likert scale with scores ranging from 0 to 72. Low scores indicate a tendency to attribute events to external factors, whereas higher scores (> 40) suggest a tendency to attribute events to internal factors. The reliability of the scale is good: internal consistency of the scale is $\alpha = .79$ and the test re-test reliability is .87 (Rallings & Webster, 2001).

The Openness to men and women (Underhill, Wakeling, Mann & Webster, 2008) scales examine the extent to which individuals are able to establish intimate relationships with men and women. Both scales consist of nine items and the response scale is a 5-point likert scale. The scales have good reliability: internal consistency, $\alpha = .85$ (Openness to Men) and $\alpha = .86$ (Openness to Women); test-retest reliability is .86 and .81 for each subscale (Underhill et al., 2008).

The Relapse Prevention Interview (Beckett, Fisher, Mann & Thornton, 1997) is an 18-item interview which elicits respondents' recognition of lapse cues, possession of coping skills and strategies, and acceptance of future risk and likelihood of relapse. Responses are coded on a 3 point scale; 0 = no recognition or skills, 1 = has some idea/skills, and 2 = shows good recognition or skills. Higher scores reflect greater relapse recognition and skills.

The Revised Dissipation Rumination scale (Caprara, 1986; Wakeling & Barnett, 2011) examines the extent to which individuals bear grudges and ruminate. It has fifteen items with yes/no responses and scores range from 0 to 30, with higher scores indicative of greater rumination. The scale has adequate reliability: internal consistency, $\alpha = .78$, and test-retest reliability, r=.64 (Wakeling & Barnett, 2011).

The Self esteem scale (Webster, Mann, Thornton & Wakeling, 2007) is a measure of general self-esteem with a yes/no response format and higher scores reflective of greater self-esteem. The scale has excellent reliability: the internal consistency is $\alpha = .84$ and the test re-test reliability is .90 (Webster et al., 2007).

The Sex Offence Attitudes Questionnaire (Hogue, 1994) is a 50-item measure comprised of six subscales relating to levels of cognitive denial, distortion and minimisation. Subscales include denial of repetition, denial of premeditation, denial of harm, denial of offence, denial of responsibility, and denial of control. Subscales are summed to produce a total minimisation score. Higher scores indicate greater subscale support. The subscales and total scale have good test retest reliability and internal consistency (Rallings & Webster, 2001).

The Sex with Children is justifiable (Mann, Webster, Wakeling & Marshall, 2007) scale examines the extent to which an individual believes that children enjoy sexual contact with adults. The scale has eighteen items, and high scores are indicative of a greater endorsement of these attitudes. The scale has excellent psychometric properties: the

internal consistency of the scale is α = .94 and the test re-test reliability is .93 (Mann et al., 2007).

The UCLA (Russell, Peplan & Cutrona, 1980) is a nineteen item measure of loneliness. The measure examines the individuals' beliefs that they had meaningful relationships, had people close to them or were lonely at the time of their offending. The response format is a 4 point likert scale, with greater scores representing a greater level of loneliness. The psychometric properties of the scale are good: the internal consistency is α = .95 and the test re-test reliability is .79 (Rallings & Webster, 2001).

The Women are deceitful scale (NOMS, unpublished) measures an offenders' beliefs about women, specifically that they are devious and manipulative. It is a 5 item measure with a 5 point likert scale response format, with higher scores indicative of the endorsement of these beliefs. The measure has good psychometric properties: internal consistency, α = .79, and test re-test reliability, r=.81. (Rallings & Webster, 2001).

New Battery Measures

The Children and sex questionnaire (Beckett, 1987) is an 87-item questionnaire that measures respondents' attitudes, feelings and thoughts about children and sex. Higher scores reflect a greater degree of attitudes supporting the sexual abuse of children. Respondents rate each item on a 5-point likert scale. Only 30 of the 87 items are scored. These 30 items are clustered into two subscales, Cognitive Distortions and Emotional

Congruence. Items are summed to produce a total scale score. High scores reflect a higher congruence and identification with children. Beech, Fisher, and Beckett (1999) report good psychometric properties for this scale, test re-test .77 (Cognitive Distortions) and .63 (Emotional Congruence).

The Emotion Control Questionnaire (Roger & Najarian, 1989) is a 56-item scale, which measures emotional response style. Respondents are asked to indicate on a dichotomous true/false scale how they feel about the 56 statements. Items are scored true = 1 and false = 2. The ECQ comprises four subscales, Rehearsal, Emotional Inhibition, Aggression Control and Benign Control. Subscales range from 0 – 28. The internal consistency of the subscales are $\alpha = .86$, $\alpha = .77$, $\alpha = .81$ and $\alpha = .79$ respectively. The test re-test reliability of the subscales are .80, .79, .73 and .92 respectively (Rallings & Webster, 2001).

The Hypermasculinity Inventory (Mosher & Sirken, 1984) is a 30-item forced-choice questionnaire developed to measure a macho personality constellation consisting of three components, Calloused sex attitudes towards women, Violence as manly, and Danger as exciting. These three components reflect the macho man's desire to appear powerful and to be dominant in interactions with others. For each item there are two possible responses, and respondents are required to choose one statement for each question. Respondents score two points for every adversarial attitude they endorse. The internal consistency of the subscales are $\alpha = .79$, $\alpha = .79$ and $\alpha = .71$ for Calloused sex, Violence and Danger subscales respectively (Rallings & Webster, 2001).

The Multiphasic Sex Inventory (MSI: Nichols & Molinder, 1984) consists of 300 items measuring individuals beliefs, attitudes and behaviours surrounding sex. It has been specifically designed for use with sexual offenders. The scale has 20 subscales including a variety of measures of sexual deviance, a measure of sexual knowledge, a measure of sexual dysfunction and a number of validity scales. The response format is true/false. Nichols and Molinder (1984) report good internal consistency: $\alpha = .89$, $\alpha = .90$, $\alpha = .85$, $\alpha = .80$, $\alpha = 94$, and $\alpha = .87$ for Sexual Obsessions, Lie, Cognitive Distortions and Immaturity, Justifications, Child Molest and Rape scales respectively. Nichols and Molinder (1984) reported test re-test correlations of .88, .92, .85, .92, .92 & .96 for Sexual Obsessions, Lie, Cognitive Distortions and Immaturity, Justifications, Child Molest and Rape scales respectively. For the purpose of this study only the MSI Sexual Obsessions and the MSI Paraphilias subscales were used.

The Relationship style questionnaire (RSQ: Dutton, Saunders, Starzomski & Bartholomew, 1994) is a 30-item self-report questionnaire measuring four attachment patterns, Secure, Fearful, Pre-occupied, and Dismissing. Respondents rate on a 5-point likert scale how well each item fits their style in close relationships. Scale anchors are: 1 = Not at all like me, 3 = somewhat like me, 5 = very much like me. Summing the items representing each subscale derives RSQ scores for the four attachment styles. Internal consistency is only available for Secure $\alpha = .41$ and Dismissing $\alpha = .41$ attachment patterns (Dutton et al., 1994).

The Social Problem Solving Inventory – Revised (SPSI-R: D'Zurilla, Nezu & Maydeu-Olivares, 2002) is a 52-item self-report measure which examines people's ability to resolve problems in everyday life. The measure consists of five subscales, Positive Problem Orientation, Negative Problem Orientation, Rational Problem Solving, Impulsivity / Carelessness Style and Avoidance style. It has a 5-point likert scale response format, with the scale anchors ranging from 0 = Not at all true of me, to 4 = Extremely true of me. Missing items are given a score of 2 (scale mean). Scores are produced for each subscale and a total score is also produced providing an overall indicator of an individual's ability in social problem-solving. Wakeling (2007) reported adequate internal consistency of the subscales PPO .73, NPO .92, RPS .95, ICS .91 and AC .82, and good test-retest reliability at .94 for the total scale.

The Paulhus scale (BIDR: 6, Paulhus 1984) is self-report questionnaire, containing two subscales, Self-Deceptive Enhancement and Impression Management. The self-deceptive enhancement subscale assesses an unconscious favourability bias closely related to narcissism. High scores on this subscale indicate the presence of a trait-like tendency toward presenting oneself in an overly self-favourable manner. The impression management subscale measures responding that is led by a desire to create a favourable impression on others and thus intends to measure the extent to which the respondent is faking or lying. High scores suggest that the individual may be exaggerating or purposely trying to impress others. Paulhus (1998) reports good psychometric properties for the measure.

Appendix C: Chapter Three Methodology

Additional Note about Methodology for Chapter Three

Prognostic modelling techniques were chosen as the preferred method here as they provide an advanced methodology for developing and validating prognostic models, or estimates of the probability of a particular outcome based on as few variables as possible. One of the aims of this chapter is to identify which variables are the best predictors of outcome (in an attempt to reduce the number of psychometrics currently used by NOMS for efficiency purposes). Prognostic modelling allows for this whilst also enabling identification of which individuals might to on to recidivate, which also has potential clinical use.

Selected Psychometrics

The current chapter examines a smaller set of psychometrics than those examined in chapter two. This decision was twofold. First, prognostic modelling techniques have particular rules for how many variables can be entered into the model according to sample size and recidivism rates which had to be adhered to. Second, some of the psychometric variables within the NOMS SOTP psychometric battery have limited predictive validity (as found from analyses conducted in chapter two). Thus, a methodology for variable selection was employed, which involved examination of all of the available psychometric scores and evidence of their predictive validity from previous studies (and from analyses conducted in chapter two), as well as expert ratings of the

usefulness of each of the psychometric variables. The procedure and results of this process are shown below.

Selecting Psychometric Variables for Models

A range of methods were used to select the psychometric variables to include in the statistical models.

1. Examination of theoretical evidence of risk factors of sexual recidivism. Which measures are dealing with risk factors which are known to be related to recidivism?
2. Examination of previous research using the psychometric variables. Has previous research shown the measures to have predictive validity? Are the measures reliable and valid?
3. Consultation with two experts. Which measures do they believe would be useful in predicting recidivism outcome?

The table below shows the outcome of this investigation.

- First, psychometrics are given a score of 0, 1 or 2 based on the degree of theoretical evidence of the risk factor they intend to be measuring being related to sexual recidivism. 0 is no evidence, 1 is some evidence, 2 is presence of good evidence.
- Second, psychometrics are given a score of 0, 1 or 2 based on the degree of research evidence existing for their validity. 0 is no research, 1 is some research, 2 is research showing predictive validity.
- Third, psychometrics are given a rating of usefulness by two experts. They both gave a rating of between 1 and 5 for each measure, 5 being the most useful in predicting recidivism. These were added together to produce a score out of 10. Then 0 points were given for scores of between 1 and 3, 1 point for scores ranging from 4 to 6, and finally 2 points were given for scores ranging from 7 to 10.
- A total of these three indicators is provided for each measure (scores range from 0 to 6).

Models were then selected based on a number of factors. Initial models were conducted using the whole sample and using measures which were gathered on the whole sample. Numbers of variables to enter into the models were based on the number of sexual and/or violent recidivists / 10.

Measures with the highest scores and which were available were entered into the model first. Offence specific measures were left out of these models. If there was a choice between two measures those which covered the breadth of risk factors known to be related to sexual recidivism (i.e. covering all SARN domains if possible) were selected.

Later models were conducted on offenders with adult victims, and offenders with child victims.

Models were also conducted using a reduced sample for which a larger number of measures were gathered (including the MSI). These models were restricted in number of variables to enter, as the sample size significantly reduces. Models were attempted for the whole sample, offenders with adult victims, and offenders with child victims.

Domain measure	**Psychometric Assessment**	**Theoretical Evidence of risk factor**	**Research evidence for psychometric**	**Total rating of usefulness**	**Total**
1	MSI – sexual obsessions (Nichols & Molinder, 1984)	2	2	2	6
1	MSI – paraphilia subscales (Nichols & Molinder, 1984)	2	2	2	6
2	Children and Sex Questionnaire - Cognitive distortions (Beckett, 1987)	2	2	2	6
2	Sex with children (Marshall, 1995 ; Mann et al., 2007)	2	1	2	5
2	MSI – cognitive distortions/immaturity (Nichols & Molinder, 1984)	1	1	1	3
2	Entitlement to sex (Hanson et al, 1994)	0	0	2	2
2	MSI – justifications (Nichols & Molinder, 1984)	0	1	1	2

2	Women are deceitful (OBPU, 1995)	0	0	1	1
2	Hypermasculinity Inventory - Callous attitudes (Mosher & Sirken, 1984)	0	0	1	1
2	Hypermasculinity Inventory - violence is manly (Mosher & Sirken, 1984)	0	0	1	1
3	Self esteem (Webster et al., 2006)	2	2	2	6
3	Children and Sex Questionnaire - Emotional congruence (Beckett, 1987)	2	1	2	5
3	Ruminations (Caprara, 1986)	2	1	2	5
3	UCLA emotional loneliness (Russell et al., 1980)	2	0	2	4
3	Locus of Control (Levenson, 1974)	0	0	2	2
3	Relationship styles – fearful (Dutton et al., 1994)	1	0	1	2
3	Relationship styles – pre-occupied (Dutton et al., 1994)	1	0	1	2
3	Relationship styles – dismissing (Dutton et al., 1994)	1	0	1	2
3	Relationship styles – secure (Dutton et	1	0	1	2

	al., 1994)				
3	Openness to Men (Underhill et al., 2008)	0	0	1	1
3	Openness to Women (Underhill et al., 2008)	0	0	1	1
4	Impulsivity (Eysenck, 1978)	2	0	2	4
4	Social problem solving inventory (D'Zurilla et al., 2002)	2	1	1	4
4	Interpersonal Reactivity Index –personal distress (Davis, 1980)	0	0	2	2
4	Emotion Control Questionnaire – Aggression Control (Roger & Najarian, 1989)	1	0	1	2
4	Emotion Control Questionnaire – Rehearsal (Roger & Najarian, 1989)	1	0	1	2
4	Emotion Control Questionnaire - Emotional Inhibition (Roger & Najarian, 1989)	0	0	1	1
4	Emotion Control Questionnaire –Benign Control (Roger & Najarian, 1989)	0	0	1	1
4	Hypermasculinity Inventory - danger as	0	0	1	1

	exciting (Mosher & Sirken, 1984)				
	Sex Offence Attitudes Questionnaire – denial (Hogue, 1994)	0	0	1	1
	Paulhus social desirable responding (Paulhus, 1984)	0	0	1	1
	Relapse prevention – coping (Beckett et al., 1997)	0	0	0	0
	Relapse prevention - recognition (Beckett et al., 1997)	0	0	0	0
	Interpersonal Reactivity Index - empathic concern, (Davis, 1980)	0	0	0	0
	Interpersonal Reactivity Index - perspective taking (Davis, 1980)	0	0	0	0
	Interpersonal Reactivity Index – fantasy (Davis, 1980)	0	0	0	0

Further Analyses Conducted Not Included in Published Version of Study

Absolute Risk Estimations

Risk estimations were also calculated based on the final three factor model for sexual offenders with different values for the three variables.

The effect of age at discharge, number of sexual appearances, and number of criminal appearances was examined based on the final three factor model using a sub sample of sexual offenders with five year follow up data (N = 1719). Absolute risk estimations were calculated for various individuals with different values for the three variables, as shown in the table below. This table reports on a selection of 30 year old and 40 year old offenders with different numbers of sexual and criminal appearances. As can be seen, the survival rates from these risk estimates are higher for those offenders who have fewer sexual and criminal appearances and who are older at the time of discharge. The lowest survival rates are associated with a 30 year old offender who has a number of sexual and criminal appearances (offender 9 below).

Table A.1 Absolute risk estimations based on the final prognostic model

Offender	No. Sex Appearances	No. Criminal Appearances	Age at discharge	Survival
1	1	1	30	0.92
2	1	1	40	0.96
3	1	4	30	0.89
4	1	4	40	0.90
5	2	2	30	0.82
6	2	2	40	0.89
7	1	7	30	0.78
8	1	7	40	0.85
9	2	7	30	0.61
10	3	5	40	0.74

Odds of recidivism were also produced from conducting a further logistic regression based on those in the sample with five year follow up data. This allowed me to calculate the probability of reconviction at 5 years for different individuals. The model was fitted entering in the three variables significant in the prognostic model (number of sex appearances, number of criminal appearances, and age at discharge). The overall model was significant ($\chi^2 = 177.48$, $p < .001$), and each of the variables were significant predictors of sexual and violent recidivism (age: Wald = 49.66, $p < .001$, Exp (B) = .95; sexual appearances: Wald = 38.53, $p < .001$, Exp (B) = 2.34; criminal appearances: Wald

= 37.02, $p < .001$, Exp (B) = 1.75; constant: Wald = 9.99, $p < .005$, Exp (B) = .39). Based on this model, the probability of reconviction for different types of offenders can be calculated. For example, a 30 year old with one sexual and one criminal appearance has a .35 chance of reconviction, whereas a 30 year old with five sexual and five criminal appearances has a .99 chance of reconviction.

These types of methods could provide valuable information to clinicians working with sexual offenders regarding the odds of recidivism for their individual clients.

Appendix D: Chapter Four Methodology

Additional Note about Methodology for Chapter Four

This chapter focuses on the importance of measuring within treatment outcome, or measuring change made as a result of treatment. As described in chapter one, it is critical to be able to measure treatment change so as to determine who may be more likely to benefit from treatment, and to help with evaluating treatment programmes. In order to be useful predictive validity indicators in this way, it is important that psychometric measures can detect the impact of treatment.

Raw difference scores were not used for the present study, as raw change scores may not fully account for where an individual starts from. For example, it has been suggested that those who are most deviant pre-treatment have the greatest capacity to evidence change (e.g. Beggs, 2010). Beggs (2010) has recommended the use of clinical significant change methodology, or controlling for pre-treatment scores when examining psychometric treatment change. Therefore Clinically Significant Change Methodology was employed in the present study, as this takes into account where someone finishes (post-treatment level of functioning), how much change they make as a result of treatment, and whether this change is reliable.

Selected Psychometrics

This study did not utilise all of the psychometrics outlined in chapter two. The same psychometrics as used in chapter three were utilised for this study. However in addition,

the MSI subscales Sexual Obsessions and Paraphilias were used as it was felt important to incorporate a measure of sexual interest into the design of this treatment change study.

Appendix E: Non-significant Cox Regression Analyses from Chapter Six

A series of Cox Regressions were performed in chapter six to examine the predictive validity of the new set of psychometrics produced via factor analyses. The tables below show the regression analyses which were non-significant. These are referred to in the main body of chapter six but the results tables are presented here for the interested reader.

Table A.2 Summary of Cox Regression examining Incremental Validity of Pre-Treatment Scores predicting General Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: General Recidivism					
Step 1: OGRS	4.71	0.52	82.04	<0.001	110.92
Step 2: OGRS	4.45	0.62	52.18	<0.001	85.88
Pre-Treatment Attitudes Score	0.01	0.01	0.46	0.50	1.01
Pre-Treatment Relationships Score	0.00	0.01	0.00	0.99	1.00
Pre-Treatment Self Management Score	0.01	0.01	0.97	0.32	1.01

Table A.3 Summary of Cox Regression examining Incremental Validity of Pre-Treatment Scores predicting Sexual Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: Sexual Recidivism					
Pre-Treatment Attitudes Score	-0.01	0.02	0.01	0.92	0.99
Pre-Treatment Relationships Score	0.02	0.02	1.46	0.27	1.02
Pre-Treatment Self Management Score	0.03	0.04	0.73	0.39	1.04

N.B. The RM2000/s did not remain in the model with the three psychometric variables.

Table A.4 Summary of Cox Regression examining Incremental Validity of Post-Treatment Scores predicting General Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: General Recidivism					
Step 1: OGRS	4.51	0.54	70.96	<0.001	90.92
Step 2: OGRS	4.29	0.57	56.06	<0.001	72.67
Post-Treatment Attitudes Score	0.01	0.01	1.53	0.22	1.01
Post-Treatment Relationships Score	-0.00	0.01	0.04	0.85	1.00
Post-Treatment Self Management Score	0.02	0.01	1.98	0.16	1.02

Table A.5 Summary of Cox Regression examining Incremental Validity of Post-Treatment Scores predicting Sexual Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: Sexual Recidivism					
Post-Treatment Attitudes Score	0.00	0.03	0.01	0.94	1.00
Post-Treatment Relationships Score	0.02	0.02	0.74	0.39	1.02
Post-Treatment Self Management Score	0.00	0.05	0.00	0.99	1.00

N.B. The RM2000/s did not remain in the model with the three psychometric variables.

Table A.6 Summary of Cox Regression examining Incremental Validity of Post-Treatment Scores predicting Sexual and/or Violent Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: Sexual and/or Violent Recidivism					
Step 1: RM2000/s	0.62	0.24	6.91	<0.01	1.86
Step 2: RM2000/s	0.55	0.24	5.17	<0.05	1.73
Post-Treatment Attitudes Score	0.00	0.02	0.01	0.94	1.00
Post-Treatment Relationships Score	0.01	0.01	1.10	0.29	1.01
Post-Treatment Self Management Score	0.04	0.03	2.35	0.13	1.04

Table A.7 Summary of Cox Regression examining Incremental Validity of Pre-Treatment and Post-Treatment Total Scores predicting Sexual Reconviction

	B	*SE*	*Wald*	*P*	*Exp (B)*
Model: Sexual Recidivism					
Pre-Treatment Total Score	0.02	0.02	1.26	0.26	1.02
Post-Treatment Total Score	0.00	0.02	0.03	0.85	1.00

N.B. The RM2000/s did not remain in the model with the three psychometric variables.

Table A.8 Summary of Cox Regression examining Incremental Validity of Pre-Treatment and Post-Treatment Total Scores predicting Sexual and/or Violent Reconviction

	B	*SE*	*Wald*	*p*	*Exp (B)*
Model: Sexual and/or Violent Recidivism					
Step 1: RM2000/s	0.58	0.29	4.02	<0.05	1.78
Step 2: RM2000/s	0.54	0.29	3.51	0.06	1.71
Pre-Treatment Total Score	0.00	0.01	0.00	0.96	1.00
Post-Treatment Total Score	0.02	0.01	2.96	0.09	1.02