ASSESSMENT OF TREATMENT CHANGE FOR SEXUAL

OFFENDERS AGAINST CHILDREN:

COMPARING DIFFERENT METHODOLOGIES BASED ON

PSYCHOMETRIC SELF-REPORT

A thesis submitted in partial fulfilment

of the requirements for the

Degree of Master of Science in Psychology

at the University of Canterbury

by

Elizabeth Anne Scott

Randolph Grace   **Anthony McLean**

Primary Supervisor   Secondary Supervisor

University of Canterbury

2015

Table of Contents

[Type here]

Table of Figures

[Type here]

Acknowledgements

I would like to thank a number of people who have made it possible for me to complete this work. My supervisor Randolph Grace provided constant support and guidance and his confidence in my abilities was invaluable throughout. I would also like to thank my employer, STOP, for their support throughout my undertaking of this research. Lastly, I would like to thank my husband, sons and all my family and friends for their overwhelming support, patience and understanding. This work would not have been possible without them.

Abstract

Communities seek success when it comes to preventing the sexual abuse of children. Thus, how best to measure treatment gains for incarcerated offenders and how those gains are linked to reductions in recidivism are important topics for research. This study examines the relationship between psychometric changes and recidivism in a sample of 495 sex offenders who completed treatment in the prison-based Kia Marama treatment programme in Rolleston prison, New Zealand. The specific goals of this study were threefold. Firstly; to characterise offender progress overall on the administered psychometric battery in terms of five different methods of calculating change. Two methods of calculating clinically significant change were employed. Firstly, change was calculated using the Jacobson and Truax (1991) method of establishing a cut off score based on normative data for each measure. Secondly change was defined as clinically significant when the post treatment score fell 1SD away from the pre-treatment mean in the direction of functionality, a methodology used by Wakeling, Beech, and Freemantle (2013). Two methods of calculating reliable change were then employed. Firstly the Jacobson, Follette, Revenstorf, et al. (1984) calculation adopted by Wakeling et al. (2013) was applied, followed by the more stringent formula proposed by Christensen and Mendoza (1986) which takes account of the standard error of difference. Finally, residual change scores were calculated, replicating the change methodology adopted by Beggs and Grace (2011). The second goal of this study was to compare the five different methodologies above for assessing change based on participants' scores on the administered psychometric battery. The third and final goal was to determine which of the five identified methods of measuring change demonstrated the strongest correlation with recidivism. Measures of clinically significant change were found to be significantly correlated with recidivism. However, this was not necessarily true when change was defined as both reliable and clinically significant. Results indicated that the Wakeling et al. (2013) method of calculating

2

clinically significant change outperformed all of the others in regards to predicting

recidivism. Overall, the present results support the use of self-report psychometrics in

measuring treatment change and predicting recidivism.

Chapter 1: Introduction

Sexual abuse: A global problem

The sexual abuse of children is recognised as a major problem globally, and the serious and traumatic impact of sexual abuse on victims is widely recognised and reported. Justified public concerns are understandably raised when the offence in question is deemed to have been predictable, and thus might have been prevented. Predictable sexual offending could be described as occurring when an offender known to the authorities reoffends in the community. Understandably, child protection agencies the world over seek to minimise this to as greater extent as possible. Clinicians who assess and treat sexual offenders within specified treatment programmes are thereby expected to provide accurate risk assessments which can contribute to the potential prevention of future offences. The efficacy of such programmes, and the accuracy of post-treatment risk assessments, are therefore crucial in the ongoing global campaign against the sexual abuse of children.

Finkelhor (1994) reported findings from a review of international data which revealed sexual abuse histories in at least 7% of females and 3% of males in the general population. Baker and Duncan (1985) reported findings of 10% for Great Britain whilst Mendelson and Letourneau (2015) state that approximately 9.3% of reported child maltreatment cases in the U.S. in 2012 involved sexual abuse. Sadly, recent evidence does not indicate a decline in prevalence, with Finkelhor, Shattuck, Turner, and Hamby (2014) reporting that 26.6% of girls and 5.1% of boys in the U.S. have experience sexual abuse by the age of 17. Fanslow, Robinson, Crengle, and Perese (2007) studied prevalence rates of child sexual abuse in women in New Zealand and reported rates of 23.5% from urban regions and 28.3% from rural regions with the majority of perpetrators being male family members of the victim. In a meta-analysis of prevalence of child sexual abuse internationally, Stoltenborgh, van IJzendoorn, Euser, and Bakermans-Kranenburg (2011) reported a consistent finding of higher

prevalence among girls than boys. Baker and Duncan (1985) reported similar results including lower ages of victimization for girls than boys They also estimated that there were over 4.5 million adults in Great Britain who had suffered sexual abuse as a child. Underreporting of child sexual abuse is common and makes the current identified prevalence rates all the more concerning (Stoltenborgh et al., 2011). There are many reasons for such underreporting, including victims often being fearful that they will not be believed, feeling too ashamed to disclose the abuse that they have suffered or worrying that they themselves are to blame in some way. As such, the actual scope of abuse may differ sharply from the officially estimated prevalence.

Alternative statistics methods of capturing the prevalence of sexual offences involve identifying recidivism rates for known offenders, that is, the percentage of sexual offenders who are reconvicted for a new sexual offence. The observed rate for sexual recidivism reported in an early, widely-cited, meta-analysis is approximately 10% to 15% after 5 years (Hanson & Bussiere, 1998). With more recent studies suggesting that the recidivism rate is decreasing (e.g. Langan, Schmitt, & Durose, 2003; Przybylski, 2015). Compared with recidivism rates for other crimes, sexual offence recidivism is lower (Beck & Shipley, 1989; Hanson & Thornton, 2000) however Beech, Mandeville-Norden, and Goodwill (2012) report statistics estimating the actual rate of sexual recidivism to be about 4 to 5 times that of the official recidivism rate. The seriousness of sexual crimes and the potential for victim harm, thus clearly warrants the attention from research, justice departments and policy makers.

Substantial research has therefore been completed in regards to the overall issue of sexual offending against children. The focus of such work has varied widely including seeking to explain why individuals engage in such behaviour, factors common to sexual offenders and how society may seek to protect communities from the perpetrators of such crimes. In depth study has been conducted to identify in to the key characteristics of sexual

offenders (e.g.Hanson & Morton-Bourgon, 2005) with Cognitive distortions, relationship and intimacy deficits and deviant sexual preferences, among others, having received particular attention (Marshall, Marshall, & Ware, 2009; Serin, Mailloux, & Malcolm, 2001; Ward, Hudson, Marshall, & Siegert, 1995). This has led to the development of numerous approaches relating such factors to actual risk of reoffending. In turn, risk assessment tools have been developed to aid practitioners in their attempts to predict (and thereby ultimately prevent) recidivism (e.g. Beech, Fisher, & Thornton, 2003; Bonta, 1996; Hanson, Harris, Scott, & Helmus, 2007; Hanson & Thornton, 2000) .

One aspect of society's response to the realities of sexual abuse has been the development of numerous psychological treatment programmes which aim to reduce the likelihood of recidivism. However, there are significant challenges to implementing such programmes. Within community programmes, generalized stigma, understandable fear, varied public opinion and, specifically, negative media responses can place significant barriers on agencies attempting to inaugurate treatment centres. One definitive example is the closure of the Wolvercote Clinic, a residential treatment centre for child sex offenders in the UK. The clinic was well regarded as being successful in the treatment of offenders, however it was closed in July 2002 following public protests and media pressures which prevented them from moving to new previously secured premises (Matravers, 2013). Other potential challenges faced by treatment providers include the impact on the clinicians employed to deliver treatment. The role itself requires clinicians to be exposed to sexually abusive information disclosed by offenders in treatment, alongside details of other child protection concerns, deviant sexual interests and pornography. In a study exploring the impact on clinicians specifically within New Zealand, Slater and Lambie (2010) acknowledged the paucity of research in this area and highlight the focus within the literature of negative impact on sex offender treatment professionals. However, they conclude that, despite therapists

facing a number of challenges, these were counterbalanced by significant rewards from the role.

The most important question regarding the treatment of sexual offenders is whether such programmes work. Over time, many studies have investigated the effectiveness of treatment programmes and meta-analyses have found reasons for optimism (e.g.Hanson et al., 2002; Lösel & Schmucker, 2005). Whilst the methodological quality of studies included in the meta-analyses has been criticised (e.g. Harris & Rice, 2003) overall the field appears to be leaning towards optimism in regards to the value of delivering treatment for sexual offenders. However, research continues on the measurement of treatment effectiveness and outcomes.

This review seeks to provide the reader with an overall understanding of how the assessment, treatment and management of sex offenders has developed over time. The development of treatment approaches is considered first, exploring early behavioural interventions through to modern day cognitive and relapse prevention frameworks from which specific models have emerged. Findings regarding the efficacy of such programmes will be presented, alongside methodological difficulties commonly faced by such studies. An overview of the development of risk assessment strategies and frameworks will also be discussed, providing the reader with a summary of current approaches. Treatment programmes in New Zealand will be described to provide a context for the present study.

The review then considers outcome research and how progress in treatment is assessed. Specifically, the construct of Clinically Significant Change and the Reliable Change Index (Jacobson & Truax, 1991) are reviewed, which provide the topic for the current study. A history of the concept and subsequent development and contextualization is presented, alongside specific applications within the field of sex offender treatment. Relevant and recent studies applying this and other methods of measuring change are presented. As will be shown, results of these studies have been mixed, which may be due in part to the use of

different methods of assessing treatment change. Thus the aim of the present study is to compare different methodologies for measuring treatment change within a single sample of treated sex offenders. Hopefully results will provide a clear picture of the relative utility of different methods for assessing treatment progress based on change scores and the power of these methods to increase predictive validity for recidivism.

The development of sex offender treatment: How things have changed

As previously noted, humanity has a long standing history of managing the prevalence of sexual abuse. With a majority of convicted sexual offenders ultimately returning to live within our communities, unsurprisingly for some time, attention has been given to seeking out ways to reduce recidivism. Early interventions focused on physical and/or biological methods such as chemical castration, aversion therapy and hormone altering therapy which sought to alter deviant sexual arousal. One example is androgen deprivation therapy (ADT) which suppress the production and action of male hormones, specifically testosterone. Androgen deprivation can be achieved both surgically and pharmacologically. Such approaches were primarily based on the assumption that deviant arousal was the root cause of sexual offending (Kirsch & Becker, 2006). However, ultimately, limited evidence was found to confirm a causal relationship between deviant arousal and sexual offending and behavioural approaches were considered to have only short term benefits. Furthermore, questions were posed around the ethics of such approaches and concerns were raised around human rights (Rice & Harris, 2011).

As understanding of the etiology of offending increased, approaches to treatment changed. By the 1970's and 1980's cognitive processes were considered to play a more central role as a contributor to sexual offending and new interventions were developed as a result. A lack of social skills was proposed by Marshall (1971) to be an important factor in the etiology of sexual offending. Subsequently interventions began to include approaches to

enhance assertiveness and other factors which were likely to contribute to developing and maintaining healthy adult intimate relationships, thereby seeking to decrease interest in deviant relationships. Similarly other patterns of thinking regarding etiology began to emerge regarding cognitive distortions and the ways in which these contributed to the justification and minimization of sexual offending. Abel, Mittelman, and Becker (1985) recommended incorporating the challenging of such distortions in to treatment interventions in order to reduce future offending.

As a result of these developments, cognitive behavioural approaches to sex offender treatment became common. Cognitive Behavioural Therapy (CBT) is one of the most popular approaches to psychotherapy and is widely considered to be the most effective method of treatment for a wide range of psychological disorders and symptoms. CBT is based on the principle that thoughts attitudes and beliefs govern our feelings and behaviour. Therefore changes in our cognitions can engender differing outcomes in terms of our emotions and subsequently our behaviour. The overall goal of treatment programmes for sex offenders is a reduction in recidivism which in itself is a change in behaviour. It is plausible to conclude that a CBT approach will be successful for treatment of sexual offenders, specifically as the etiology of sexual offending is considered to be multi-causal and therefore a comprehensive approach to treatment is required (Becker & Murphy, 1998).

There are clear reasons why CBT should be appropriate for sex offender treatment. Therapeutic interventions for sex offenders are based around a central premise that changes in an individual's behaviour are possible. This is why dynamic risk factors have been identified as different and separate from static factors, because of their potential for change. Hence, in using a CBT approach, many of the key dynamic factors associated with recidivism can be targeted. An example of this includes the cognitions which justify, minimize and maintain sexual offending behaviour can identified and challenged in order to reduce future

risk or reoffending. Therefore modules focusing on cognitive distortions are a common part of treatment programme content. Cognitive behavioural therapy is now considered to be the most common and consistently effective approach for sex offender treatment programmes (Becker & Murphy, 1998; Marshall, Eccles, & Barbaree, 1993; Moster, Wnuk, & Jeglic, 2008).

With CBT appearing to be the method of choice to help sex offenders understand their offending and change their thinking, the inclusion of relapse prevention frameworks began to be included in treatment programmes to enable individuals to maintain changes, thereby reducing recidivism (Moster et al., 2008). Relapse prevention (RP) was initially designed as a maintenance programme for individuals who attended rehabilitation programmes for drug and alcohol abuse. According to the RP model, individuals must identify the key triggers likely to contribute to relapse and develop coping strategies and interpersonal skills to enable them to manage the factors that increase the risk of relapse occurring. Whilst not necessarily directly identified in the etiology of sexual offending, the RP approach was rapidly identified as having significant potential when applied to sex offenders and became a key component of many treatment programmes (Moster et al., 2008; Pithers, Marques, Gibat, & Marlatt, 1983). Whilst the initial framework for relapse prevention was developed in the 1980's (Marlatt & George, 1984) it was later modified by Pithers (1990) for specific application to sex offenders. Hudson and Ward (1996) later identified problems with Pither's approach and proposed a further adaption which more sufficiently took into account the offence process. This is now commonly addressed in most treatment programmes in modules focusing on understanding patterns of offending, often referred to as 'offence chains'. Kirsch and Becker (2006) comment that, whilst studies analysing the efficacy of the incorporation elements central to the RP model are limited, RP continues to be included in the framework for cognitive behavioural therapy used in a majority of sex offender treatment programmes.

In their longitudinal study of sex offender treatment in California, Marques, Wiederanders, Day, Nelson, and van Ommeren (2005) did not find results which supported the efficacy of an Cognitive behavioural RP approach. However they comment on a number of methodological issues in their research and specifics of their treatment programme which could account for this. One of the key issues identified is the concept of their programme not fully meeting all three of the principles of the Risk-Need-Responsivity model (RNR), first formalized by Andrews, Bonta, and Hoge (1990). This model is based on the concept of effective rehabilitation (effective being deemed achieved with observed reduction in recidivism) needing to incorporate the principles of risk of recidivism, criminogenic need and responsivity of offenders in their assessment and treatment framework. The model was further contextualized and elaborated on with the aim of enhancing the design and implementation of treatment programmes and soon became a key aspect of the literature on 'what works' with sexual offenders. Hanson, Bourgon, Helmus, and Hodgson (2009) conducted a meta-analysis of 23 studies examining recidivism outcomes. They concluded that programmes which incorporated RNR principles demonstrated the largest reduction in recidivism, both sexual and general. Such findings offer clear support for a model which has understandably become associated with best practice approaches to sex offender treatment.

Given the ever expanding nature of research in to assessment, treatment and management of sex offenders, unsurprisingly, the RNR model was not without its challengers. There appeared to be a general consensus around its particular usefulness and success in its aims, however critiques and reviews arose. Ward and Stewart (2003) noted the overall importance of seeking a balance within risk management and considering the overall needs of an offender. They highlighted the importance of satisfactory lives for offenders and human well-being for individuals engaging in treatment. They compared this to the RNR model which they perceived to place greater importance on risk management avoiding harm

to the community than on the quality of life of the offender, that is, focusing on avoidance rather than approach methods of managing risk. Ward and Stewart (2003) concluded that, by enhancing an individual's capacity to meet his needs in adaptive ways, he would be less likely to seek out ways to harm others, thereby reducing recidivism and the strengths based approach of the Good Lives Model GLM was proposed. Ward and Gannon (2006) further conceptualized this model. They proposed a collaboration of the original Good Lives Model (GLM-O) (Ward & Stewart, 2003) with the Integrated Theory of Sexual Offending (ITSO) (Ward & Beech, 2006). The ITSO posited that there were a number of causal variables which could explain the onset, development and maintenance of sexual offending, including neuropsychological, ecological and clinical factors (Ward & Beech, 2006). In their drive to improve on earlier work, Ward and Gannon (2006) advocated the Good Lives Model – Comprehensive (GLM-C). They proposed that this would provide a novel, more comprehensive and broader theoretical framework on which to develop effective treatment interventions. Ward and Gannon (2006) described their model as addressing many of the areas critiqued on previous framework and deemed it to encompass all relevant areas and needs of an effective treatment approach, namely; ecological and developmental factors, a multi-systemic approach, risk management concerns, goods promotion, consideration of humanistic and scientific values, and emphasizing the importance of tailoring treatment to individual need and forward planning. Rather than being seen as competing theories however, Ward, Melser, and Yates (2007) highlighted the potential gains in a more complementary approach with the GLM supplementing the RNR in specific areas.

Despite the shift from the behavioural to the more cognitive approach in sex offender treatment, there are still frequently used behavioural components incorporated in to many treatment programmes. Commonly these include aversion therapy, covert sensitization, masturbatory reconditioning, and directed masturbation (Kirsch & Becker, 2006). Such

techniques often aim to modify sexual preferences and alter sexual arousal patterns (Laws & Marshall, 1991) and often include attempts to reduce deviant sexual arousal and increase appropriate arousal to adult stimuli . Laws and Marshall (1991) conclude that there are low numbers of controlled studies and no group comparison studies, thus making it difficult to draw definitive conclusions on the effectiveness of such techniques. However they further comment that the evidence available is promising and that directed masturbation or satiation appear to be the most promising masturbatory reconditioning techniques.

Of note, despite the ethical concerns raised decades ago regarding pharmacological approaches to treatment, currently treatment such as ADT continues to be used as an approach in the treatment of sex offenders. Nine U.S. states and some other countries (e.g. Czech Republic) have laws governing chemical or surgical castration for sex offenders whilst others have laws which permit pharmaceutical intervention as part of a treatment strategy for community based sex offenders (Rice & Harris, 2011). This is the current situation regardless of the overall paucity of empirical evidence in support of such techniques in reducing recidivism in sex offenders.

Given what is known about the heterogeneity of the pathways to sexual offending, it is unsurprising that approaches to assessing and treatment offenders has varied widely across time. Whilst there still continues to be no uniform approach across treatment programmes, it is apparent that there are significant commonalities in both the methods, modalities and models of treatment provided and research can be seen as ever expanding in this area. This was perhaps best summed up by Ward et al. (2007) (p 226):

"*Whatever occurs, it is our conviction that we need to continue evaluating, refining, and at times reconstructing our rehabilitation models. As Andrews and Bonta (2003) have frequently stated, opened-ended critical inquiry is an essential part of the correctional enterprise. We must never assume there is no more to be learned; the stakes are far too high*

to insulate ourselves from ongoing criticism and debate. Vigorous debate is the lifeblood of

science and the source of effective and humane practice"

Treatment in New Zealand

Currently a variety of modalities of treatment are available for sex offenders in New Zealand. The Corrections department operates two specialist treatment programmes based in Auckland (Te Piriti) and Christchurch (Kia Marama) which offer treatment to incarcerated child sex offenders. Three community based treatment programmes operate out of Auckland (SAFE), Wellington (WELSTOP) and Christchurch (STOP) and Community Probation Services Psychology department (CPPS) staff work with individuals serving non-custodial sentences. Alongside this, a small number of private practice clinicians offer services for sex offenders ranging from risk assessments, one to one treatment interventions and general counselling and self-management support.

Efficacy of treatment

The development of numerous psychological treatment programmes which aim to reduce incidents of recidivism has been one aspect of society's response to sexual abuse. But do such programmes work? Given the gravity of decisions made based on outcome in treatment, not to mention the financial implications of developing and providing treatment programmes, it is clearly imperative that these therapies are actually successful. With a universal goal of the reduction of offending and therefore a decrease in levels of victimisation, it is clear to see why a majority of studies have focused on recidivism data to evaluate the effectiveness of treatment programmes for sexual offenders.

Over time, many studies have investigated the effectiveness of such programmes and meta-analyses have found reasons for optimism (e.g. Lösel & Schmucker, 2005). Probably the most cited of such analyses is Hanson et al. (2002). However, the methodological quality of studies included in this meta-analyses has been criticised and general findings have been

described as varied and controversial (Beggs & Grace, 2011; Rice & Harris, 2003). Thus, whether treatment for sexual offenders against children is effective remains an unresolved question, and so continued research on the measurement of treatment effectiveness and outcomes is warranted, hence the rationale for this research.

Hanson et al. (2002) analysed 43 studies, giving them a sample size of over 9000 offenders. When averaged across studies, they found lower recidivism rates for the treated sample than untreated. They concluded that current treatment programmes (cognitive behavioural and systemic) had a positive effect on reducing recidivism but older approaches appeared to have little effect. Hanson et al. (2002), and most other studies, highlight the challenges faced by effectiveness studies in obtaining accurate and valid results. These include follow up periods, low base rates of treated and untreated offenders and small sample sizes resulting in questionable statistical power. Beech, Friendship, Erikson, and Hanson (2002) reviewed three prior meta-analysis studies seeking to assess effectiveness of sex offender treatment (Alexander, 1999; Gallagher, Wilson, Paul Hirschfield, Coggeshall, & MacKenzie, 1999; Hall, 1995). Overall some positive findings were noted in regards to overall treatment effect however Hanson et al. (2002) highlighted specific methodological considerations likely to have impacted results, These included comparing samples from different programmes, recidivism criteria, comparisons of completers and dropouts and differing time at large (see Hanson et al., 2002 for a review).

Seeking to address such methodological bias, Hanson et al. (2002) focused on subject assignment and distinction between older and current treatment programmes to analyse groups of offenders receiving psychological treatment for sexual offending (N= 5078), versus untreated offenders or those receiving treatment judged to be inadequate (N = 4376). With an average 46 month follow up period, sexual recidivism rates of 12.3% and 16.8% were found for the treated and untreated groups respectively with similar results for general recidivism.

Further findings concluded that studies comparing current treatment to alternate treatment approaches found greater treatment effects and some evidence for stronger effects for community based programmes compared with institutional based treatment. Hanson et al. (2002) recommended that their results should be interpreted cautiously, specifically in relation to the relatively low number of studies.

The large sample size used by Hanson et al. (2002) provided obvious reason for optimism, however overall scepticism of meta-analysis studies prevail within this work and their conclusions were not without their challengers. Rice and Harris (2003) raised questions around the effectiveness of the Hanson et al. (2002) paper. They draw an interesting distinction between efficacy (can treatment work?) and effectiveness (does treatment work?) and comment that most research has focused on the latter for obvious socio-political reasons. Rice and Harris (2003) provide strict criteria for what they perceive to be an essential standard in efficacy studies. They insist on comparable groups and officially recorded recidivism from at least two groups of offenders (with at least one group receiving treatment) and criticise Hanson et al. (2002) comparisons of treatment completers with offenders not offered treatment. Further criticisms include low numbers in the random assignment design study (a design noted as preferable by Rice & Harris, 2003), overreliance on incidental assignment studies and the exclusion of specific studies which notably reported adverse treatment effects. Whilst the authors acknowledge their evaluation of the literature to be "conservative" (p. 11) they offer useful observations around the gaps in the then current research regarding best practice programme content and types of offender most likely to benefit from treatment. They further highlight the need for strong inference techniques to ensure that research contributes to progress in the field of sex offender literature.

The above two studies have been widely cited in the field, with many offers of critique, comparison and support. Whilst appearing to be opposing at first glance, the

similarity of both of these studies lies in the conclusions drawn around the need for further, methodologically-sound approaches seeking to provide firm conclusions regarding efficacy and effectiveness.

Whilst findings may still be mixed in regards to the effectiveness of treatment programmes, it is essential that both the assessment and treatment of offenders continue. Society simply cannot wait for a robust, all-encompassing and unchallenged method of treatment to be developed, leaving many offenders untreated and poorly managed within the community and therefore numerous children at risk. Moreover, it is impractical and unethical to expect all offenders to be incarcerated until research is concrete about the effectiveness of treatment. Hence something must be done. It appears overall that there is a wide and growing acceptance that treatment programmes following the current findings on 'what works' with offenders can be effective in reducing recidivism and the provision of such programmes can begin to meet society's demands for a reduction in the victimization of children.

Methodological Difficulties with evaluating treatment programme data

The assessment, treatment and management of individuals at risk of sexual offending is an ongoing challenge faced by society. There is a clear role for research which can provide evidence of ways to enhance all aspects of the management of offenders. Much of this research relies on the analysis of data collected from current treatment programmes and the methodology of studies varies greatly. Whether the aim of specific studies is to explore the efficacy of a programme or treatment changes linked to recidivism is the target outcome, often researchers encounter difficulties in the analysis of the data. The systematic evaluation of data is often difficult, and this section aims to highlight some of the common methodological challenges raised in the evaluation of treatment programme data in general.

Typically self-report psychometrics are incorporated as part of most studies, here the issue of social desirability can be an important consideration. As noted previously, offenders

often have significant investment in presenting themselves in a positive light at the end of treatment (Tierney & McCabe, 2001). It is understandable that their results on post treatment, and indeed pre-treatment, measures may reflect this rather than demonstrating actual genuine change in attitudes, beliefs and behaviours (Hanson & Bussiere, 1998). The result can be an overestimation of change achieved or efficacy of programmes.

A further difficulty with the use of self-report psychometrics is the transparency of the measure itself. Linked to the issue of social desirability, a transparent measure can significantly influence the accuracy of the change perceived to be evident at post treatment. Therefore outcome data can only be considered to be as good as the measures used. The validity and reliability of a measure must therefore be considered and commented on in some way as an integral part of the study's analysis of data. Similarly, change across psychometrics is often measured based on raw difference scores. Among other issues raised here, such methodology does not take in to account the pre-treatment score and can results in an overestimation of change. The potential here is for this in turn to lead to an overestimation of efficacy should this approach be used in studies of programme effectiveness studies. As detailed throughout this review, a growing line of research seeks to address this issue and forms the main rationale for the current study.

To overcome issues related to small sample sizes and other methodological challenges, meta-analysis have often been conducted (e.g. Hanson & Bussiere, 1998). These studies seek to analyse data from a number of studies and thereby provide conclusions based on a large population of offenders. Unfortunately, whilst this resolves some of the challenges noted above, further difficulties are found. Often the overall sample of a meta-analysis is derived from many populations. Within this, there are likely to have been significant differences in a number of areas. The actual treatment received by individuals is likely to vary significantly. Treatment delivery modalities vary from programme to programme, as do

programme content and duration (Lösel & Schmucker, 2005). Meta-analysis also often faces the challenge of seeking results from studies which have drawn their conclusions based on a variety of measures used. Some studies use sample data from individual programmes, however here often difficulties are faced in regards to low base rates of offending and small sample sizes (Furby, Weinrott, & Blackshaw, 1989; Hanson & Bussiere, 1998). Significant results can therefore be difficult to achieve and are easily challenged.

Often studies evaluating treatment for sex offenders utilise recidivism data in some manner within their analysis. This in itself poses challenges. The definition of recidivism can differ from study to study. Some define recidivism as re-arrest, others use reconviction as their benchmark. The type of offence can also differ with some studies counting any sexual or violent offence, and others including only sexual recidivists (Hanson & Bussière, 1998). With a general acknowledgement currently, that actual re-offending rates are greater than recorded reconviction rates (Furby et al., 1989), relying solely on reconviction information can lead to studies missing out on important data which could be more likely to give an accurate measurement of treatment efficacy. Here the obvious question however is how more actual re-offending rates could be reliably obtained or estimated.

Further challenges linked to recidivism data lie within the complexities of follow up periods. The length of follow up varies significantly across studies and can affect the reported results. Longer follow up periods are often considered to provide greater statistical power in data analysis. However, such longer periods can be difficult to achieve within sex offender research (Hanson et al., 2002). With such importance placed on finding the most effective way to reduce recidivism, there is significant current pressure to ensure that risk assessment tools are as accurate as possible and for treatment to be maximally effective. There is therefore a sense of urgency in research. Longer follow up periods can also result in studies analysing data from particularly old treatment programmes which may not follow the current

19

trend in regards to programme content, length selection criteria etc. Hence results may be skewed by these factors rather than providing accurate data analysis (Lösel & Schmucker, 2005).

When studies include data from various samples, other discrepancies in the populations can arise in regards to treatment selection. It seems apparent that currently a majority of treatment providers follow the risk principle of Andrews and Bonta (2003) Risk-Need-Responsivity model in regards to treatment selection. Offenders are therefore categorised based on a nominal risk level assigned to them following assessment using some form of risk matrix. This generates differences across studies as a number of risk assessment tools are available to clinicians and thereby meta-analysis are likely to be managing data based on risk levels determined by different measures of risk. The decisions made based on risk are also likely to differ. For example, some programmes deem low risk individuals to be unsuitable for treatment whilst others would instead offer the offender a shorter treatment programme. Similarly, high risk individuals tend to receive treatment for a longer duration.

Alongside risk assessment, there are often differences in programme entry requirements. The inclusion and exclusion criteria used to make decisions around treatment selection again contribute to variety across samples. Intellectual functioning, active mental health concerns, current drug and alcohol dependence, psychopathy, motivation to engage in treatment and denial of offending are common factors considered by treatment providers in regards to selection. Alongside this, risk management issues, specifically in regards to immediate risk of harm and other safety concerns can be influential in selection decision making for community based programmes, whilst sentence length and accessibility to treatment centres can pose other challenges in the prison population. Not all of these decisions are based on of empirical evidence such as actuarial measures. Clinical judgement has a part to play here and therefore subjectivity of individual clinicians adds a further layer

to the challenges of methodology. Lastly, funding and contracts for individual programmes can play a role in decision making around treatment selection. Pressure can be felt by programmes in regards to meeting contracts and achieving targets for numbers in treatment. Similar issues are found within the range of completers. Attrition rates vary across programmes. Similarly, decisions in regards to exiting an offender from treatment are often based on clinical judgement and the aforementioned contractual pressures can also be influential here. All these factors serve to further complicate the make-up of any given sample when individual or meta-analysis are conducted.

Assessment of Sex Offenders

The assessment of offenders who enter the criminal justice system has many forms and functions. Assessments are initially made to determine sentence type, length and conditions and restrictions. Further assessment is usually conducted pre-treatment to determine treatment duration and modality required and identify key areas of criminogenic need for effective treatment planning. Post-treatment assessments are usually then performed to contribute to treatment outcome evaluations and decision making. The form of these assessments varies depending on the purpose and research has developed significantly in this area over the years.

<u>Risk Assessment</u>

Risk of recidivism in sex offenders is of specific interest to clinicians, policy makers and the public alike. Critical and potentially life-changing decisions are made based on predictions of risk, including type of sentencing, length of incarceration, sentence conditions and limitations regarding contact with children. Accurately predicting recidivism is therefore of utmost importance, however such a task is complex and multifaceted. Significant focus has therefore been placed on this in sex offender literature as numerous and varied approaches have been made to the development of accurate risk assessment tools. Studies have often found over classification of risk to be common among even experienced risk assessment

professionals (e.g. Lanterman, Boyle, & Ragusa-Salerno, 2014). Furthermore regularly updating risk assessments is also crucial given the extent to which an individual's circumstances may change over time. This could lead to a reduction or increase in risk and monitoring such changes is imperative for effective risk management. For example, (Hanson & Morton-Bourgon, 2004) and (Lambie & Stewart, 2012) found an increase in recidivism for individuals who attended but did not complete treatment. Accurate and regular risk assessment is therefore deemed to be crucial for the reduction of sexual offending.

Early approaches to risk assessment are now often defined as first generation risk assessment tools. These were prevalent prior to the 1990's and involved unstructured clinical judgements by professionals working with the individual offender. Easily identifiable flaws are apparent with such an unguided approach, including lack of inter-rater reliability and inconsistencies in the specific items considered. In response to this, structured clinical judgements became more widely used with more consistency placed on items specified for consideration, however these still lacked the use of any validated matrix or measure and were therefore still open to bias and personal interpretation. (Hanson & Morton-Bourgon, 2004) found that unstructured clinical assessments were significantly related to recidivism but that their accuracy was consistently lower than actuarial measures.

Second generation risk assessments therefore sought to enhance predictive validity and eliminate interpersonal bias. Empirically-validated actuarial measures were therefore developed which based risk assessment on offence-history information and thus eliminated subjective bias. (Andrews, Bonta, & Wormith, 2006) found that classic second generation instruments did well in the prediction of general recidivism. Assessment tools specific to sexual offender recidivism were therefore developed which take in to account only static, historical risk factors which are unchangeable by direct intervention for example age, victim gender and criminal history. Examples of actuarial risk assessment tools include Static 99R

22

(Hanson & Thornton, 2000), RM2000 (Thornton et al., 2003) and the ASRS (Skelton, Riley, Wales, & Vess, 2006) which was developed specifically for use with a New Zealand population. The Static 99R (Hanson & Thornton, 2000) comprises ten items, e.g. age, marital status, gender of victims and number of prior sex offences. It is the most widely-used actuarial measure for sexual offenders and has been shown to have moderate to good predictive validity for sexual recidivism across a range of evaluation studies for example (Harris & Rice, 2003).

Second generation risk assessment tools do not take in to account factors which are changeable by psychological intervention and therefore do little to inform clinicians in regards to the areas on which treatment should focus. Basing risk predictions on static tools alone implies that an offender would never be able to change his or her risk level and as such, impose limits on risk management in general. Such a proposal would nullify the need for interventions which work to reduce recidivism and thus contradict the overall goals of justice systems which seek to rehabilitate offenders and reduce criminal behaviour. Such limitations of tools which rely solely on static factors were swiftly identified and therefore third generation risk assessment measures were sought. An approach to address the limitations of second generation assessment tools was first pioneered by Andrews et al. (1990) with the formalization of their Risk-Need-Responsivity model. This has been significantly elaborated on over time (Andrews et al., 2006; Bonta, 1996) with the 'Need' principle contributing significantly to the development of tools to assess psychological and behavioural features, termed dynamic risk factors, which contribute to overall risk of recidivism and are potentially amenable to change. Examples of such factors in sex offenders include distorted attitudes, problem solving abilities and (deviant) sexual interests. Professionals are therefore able to use information about dynamic factors as a focus for the planning of the overall content and delivery of treatment programmes as well as specific individualized treatment plans for

offenders. There are many examples of third generation tools that encompass such dynamic factors and criminogenic need including STABLE 2007/ACUTE-2007 (Hanson et al., 2007), the Structured Assessment of Risk and Need (Webster et al., 2006) and the Violence Risk Scale-Sex Offender Version (M. Olver, S. Wong, T. Nicholaichuk, & A. Gordon, 2007a).

Fourth generation assessments seek to take a step further towards defining what is important for professionals to consider in regards to risk and are considered to be best practice for accurate risk assessment. This level of assessment takes in to account treatment change thus identifying changes in dynamic risk factors. Such assessments also integrate the 'responsivity' principle of Andrews et al. (2006) framework, together with static and dynamic factor assessment to produce a systematic and comprehensive approach. They define the responsivity principle as seeking to "maximize the offender's ability to learning from a rehabilitative intervention by providing cognitive behavioural treatment and tailoring the intervention to the learning style, motivation, abilities and strengths of the offender". Thus such assessments will provide a detailed and all-encompassing approach to risk assessment and management.

A further focus in risk assessment research recently has been on the need for, and importance of, a common language by which professionals can communicate risk levels. Categorical statements of risk, for example "low" or "high" risk are clearly open to individual interpretation even when clinicians are familiar with a specific actuarial scale. Risk is often communicated in statements of relative risk i.e. comparing offenders to none offenders and in absolute recidivism rates i.e. numerical rates of recidivism for a specific risk band/level (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012). Neller (2013) reported limitations of nominal risk statements and highlighted the need for numerical probability statements of risk to be included in professional dialogue. However they also emphasized that numerical probability statements based on proportions rather than predictive values could be

misleading. It is apparent this issue is just one aspect of the ever expanding field of research in regards to seeking the ultimate gold standard in sex offender risk assessment.

Psychometric Evaluation

Most treatment programmes employ some form of psychometric assessment of programme participants to collect data regarding personality characteristics, functioning and attitudes and beliefs. There are considerable differences in the specific measures used. Some measures have been developed for use within the general population but assess psychological constructs which have been found to be relevant to sex offender populations. Examples here include the Beck Depression Inventory (Beck & Ward, 1961) and the Social Self-esteem Inventory (Lawson, Marshall, & McGrath, 1979). Other scales are more offence specific and are designed to provide information specific to a sex offender population e.g. The Multiphasic Sex Inventory (Nichols & Molinder, 1984) and The Abel-Becker Cognitions Scale (Abel et al., 1989). Whilst the administration of a battery of psychometrics pre and post treatment is commonplace, opinion on their usefulness is divided. Much of the research in this area has tended to focus on psychometric scores and their link to recidivism, with mixed results.

Wilkinson (2005) analysed psychometric data from a cognitive behavioural training programme for incarcerated offenders in England and Wales. Findings indicated that offenders who demonstrated pro-social changes in attitude were more likely to be reconvicted than those who did not. The author cites possible methodological grounds for the unexpected result and questions the use of attitudinal measures in treatment outcome measurements. Similarly, Proulx et al. (1997) studied predictors of recidivism in sexual aggressors. Their sample comprised 382 adult males assessed in a maximum security psychiatric hospital using a battery comprised of a mixture of offence specific measures and general measures of social inadequacy. Results indicated that psychometric data did not predict recidivism in rapists or child molesters with the

authors drawing conclusions that potentially the constructs being measured were themselves not good predictors of recidivism. The small sample size may well be a relevant factor here also as the psychometric assessment battery was not introduced to the unit until 1987. Consequently, pre-treatment psychometric data was only available for a small proportion of the sample (N=66) with only 25 participants re-assessed subsequently post-treatment.

An early addition to the more promising band of research linking psychometric data and recidivism, examined pre-treatment psychometric scores from 140 adult males who had sexually abused children (Beech, 1998). Cluster analysis revealed two specific groups of offenders. These were subsequently labelled Cluster A "high deviancy" as their mean scores deviated significantly from the non-offender norms for the measures and Cluster B "low deviancy". Significant differences were found between the two clusters in regards to prior number of offences and victim type and risk of reconviction, lending weight to the argument that psychometric data can be useful in predicting recidivism. Walters (2006) conducted a meta-analysis of 22 studies which employed one or more of five risk-appraisal procedures and one or more self-report psychometric measures. Findings indicated that self-report measures can predict recidivism however this was only for crime-related content measures as opposed to general indices of personality. Other studies found positive results between psychometrically measured deviancy level and prediction of recidivism. Beech et al. (2002) examined pre-treatment psychometric data for child abusers about to enter community based treatment in the UK. With a six year follow up period, results indicated that the addition of psychometric measures of dynamic risk provided incremental validity to a static risk assessment. Similar results were found by Beech and Ford (2006) in a study of 51 child abusers who received treatment at a residential unit in the UK. Men assessed psychometrically as high deviance were found to be more likely to be reconvicted at 2 and 5 year follow up periods than those deemed low deviancy. Low base rates of reoffending and

small sample sizes warrant caution in the interpretation of their results. However they provide some promising results in regards to the link between psychometric assessment and recidivism. Barnett, Wakeling, Mandeville-Norden, and Rakestrow (2012) examined a much larger sample size in an attempt to remove the potential effect of lower numbers of participants. Their study examined psychometric data for 3402 sex offenders who attended community based treatment in the UK. The sample comprised both rapists and child abusers and included only those for whom pre and post-treatment scores were available. In regards to individual measures, the only pre-treatment scores shown to predict recidivism were those from the Fantasy subscale of the Interpersonal Reactivity Index (IRI ((IRI) Davis, 1980). Unexpectedly, better recognition of risky thoughts, feelings and situations (as measured post-treatment on the relapse prevention questionnaire) was predictive of reoffending. Scores relating to self-esteem were also found to be predictive of recidivism, although this was only the case for post-treatment and not scores measured pre-treatment. When grouped into dynamic risk domains, post-treatment scores were higher for recidivists than non-recidivists for the Self Management domain, however more significant results were found for the Socioaffective Functioning domain with recidivists scoring more highly at both pre and post treatment. Here the authors conclude tentatively that psychometric scores, specifically those measured pre-treatment, can be useful in predicting recidivism. Of specific interest to this study is their recommendation regarding the importance of the development in understanding of the relationship between change in treatment, that is, treatment outcome (as measured psychometrically) and re-offending.

Treatment Outcome

Regarding treatment outcome, a shift in thinking has occurred with respect to individuals who complete treatment. It is no longer considered sufficient to simply conclude whether a person has completed a programme or not. Literature shows that attendance at

treatment and/or treatment completion are not deemed to be predictive of reductions in recidivism. For example, in their study exploring the impact on recidivism of extended supervision orders for child sex offenders in New Zealand, Watson and Vess (2008) found programme attendance and completion did not predict sexual recidivism. Specifics of the actual outcome of an individual's treatment have therefore become a key focus for justice departments, policy makers and treatment providers alike. With a vast number of child protection decisions being made regarding clients who complete treatment, risk assessors, social workers, family members and sometimes victims, want to know if the person has changed and if that change is linked to a likely reduction in recidivism. The general consensus now appears to be accepting of the efficacy of treatment programmes, yet reoffending continues to occur. With individuals reoffending at different rates, and some offenders remaining offence-free, there are clear variations in the extent to which individuals benefit from the treatment they receive. This provides further relevance to the need to develop measures of individual treatment outcome and justifies an investigation in to valid measures of change. Outcome reports based on risk and recidivism probabilities are often of lesser interest to offenders themselves as they mostly tend to claim that they will never reoffend. Specific evidence regarding the extent to which they have changed and benefited from treatment is therefore of more interest to them.

Beggs and Grace (2010) provided a detailed summary of differing rationales relating to the application of assessing within-treatment outcome. The review of the literature is comprehensive and detailed and critical to this research and therefore will be summarized here. Beggs (2010) cites differing rationales and applications of the process of measuring change. For example, measurements of change contributing towards more clarification regarding the efficacy of treatment programmes by moving away from the tendency to assess completers as a single cohort and only compare treated offenders to those who refuse or drop

out of treatment (Anderson, Gibeau, & D'Amora, 1995; Scalora & Garbin, 2003; Seager, Jellicoe, & Dhaliwal, 2004). Three other key applications are also considered (Beggs, 2010):

1) Whether change on specific components in treatment is related to recidivism (Marques, 1999). Thus providing support to programme developers in regards to identifying key specific programme content

2) Using research from links between within-treatment change and recidivism to explore whether dynamic risk factors are in fact changeable and if those changes do in fact have links to recidivism

3) Using measurements of within-treatment outcome to add to responsivity research, that is, studies which explore links between within-treatment measures of change and various specific offender characteristics

With the shift from a dichotomy of 'treated' or 'untreated' based purely on attendance in a programme, comes complexities and challenges regarding how to measure outcome of treatment. The measure of efficacy of treatment mentioned earlier, clearly focusses on statistical and empirical validation of within group numbers and in itself is challenging enough to accurately measure. However identifying positive treatment outcome among individual offenders raises even further challenges. The rationales behind studies of outcome in treatment have varied, including seeking links between positive change and reduced recidivism and adding to evidence of overall efficacy of programmes. However offenders themselves and their families and support networks often also hold their own motivations for treatment outcome to be accurately assessed.

With often life-changing decisions based around treatment, it is would be imprudent not to consider the element of social desirability in individual's performance in treatment. When seeking early release from prison, lighter release conditions, and possible access to children or reintegration to the family home, "faking good" becomes a viable option for

treatment completers and poses a significant challenge to those who are tasked to assess genuine change. This is particularly an issue with self-report psychometrics which can often be transparent and therefore there is a risk of a social desirability bias in the responses. Tierney and McCabe (2001) report issues of social desirability in self-report measures of cognitive distortions and empathy specifically and recommend the use of a valid and reliable social desirability scale when assessing sex offenders pre and post treatment..

Measuring Treatment Outcome

As described earlier, static risk factors are those which essentially remain fixed, and thus are not able to inform about change in treatment. Dynamic risk factors are those which have the potential to change. Most sex offender treatment programmes therefore aim to target specific dynamic factors within the content of the programme. The notion that a person can change their thinking and behaviour is the construct for most cognitive behavioural interventions. Theoretically, therefore, one would expect that positive progress in treatment was associated with change across these dynamic factors and that such a change would ultimately result in a reduction in recidivism. Analysing change in treatment can therefore be closely linked to the process of post treatment dynamic risk assessment, however change can occur over a wider range of factors than those included in specific dynamic risk assessment tools. Therefore separate and specific measures are required to assess treatment outcome. If it is important to determine the outcome of an individual's treatment, the method by which this is measured must therefore be robust. Given the current interest in seeking more from treatment outcome research than just recidivism statistics, it is unsurprising that this is an emerging field. However, there is still a paucity of specific literature around this relating to the treatment of sex offenders which suggests the importance of developing useful and effective ways to measure change in treatment.

Typically, psychometric assessments designed to measure change on dynamic risk factors, are administered pre and post treatment and obtained scores are used to provide one line of evidence regarding progress in treatment. Despite concerns regarding the reliance on psychometric self-reports, studies have focused on links between psychometric data, treatment outcome and recidivism, with mixed results. In the general offending literature, Bowen, Gilchrist, and Beech (2008) found no relationship between change in treatment (as measured psychometrically) and reoffending for a UK domestic violence sample who attended community based treatment. Change was measured using clinically significant change and reliable change index calculations. Results were non-significant at group-level however positive and reliable change was evidenced for a number of individuals across the measures. Small sample sizes, short follow up periods and potentially inaccurate recidivism data were identified as limitations of the study. Seto and Barbaree (1999) concluded similar results in regards to a lack of relationship between treatment performance and recidivism. Using data from an institutional sex offender treatment programme, they predicted that more positive behaviour in treatment would be less likely to reoffend. With an average follow up time of 32 months and a general recidivism rate of 14.7%, results did not support the initial hypothesis. Surprisingly, individuals who scored highly for psychopathy and performed well in treatment were found to be more likely to commit a new offence. Seto and Barbaree (1999) highlight the limitations of short follow up times and of conducting the study retrospectively and suggest that a prospective study may reach different conclusions. The robustness of the methods for measuring treatment progress was also cited as a potential limitation of the study. This adds to the growing acceptance of the need for robust methods of calculating change in treatment. On review of these findings five years later, Barbaree (2005) re-examined the data with a more substantial set of recidivism information and longer follow up period. New conclusions were drawn that there was no link between treatment behaviour and

recidivism. Seager et al. (2004) found similar results. Their sample (N = 146) consisted of treatment completers and non-completers who attended sex offender treatment at a medium security prison. With a 2 year follow up period, 23% of offenders had received a new sexual/violent charge or conviction. Whilst the premise of this study was to consider the overall efficacy of the treatment programme, the results are relevant here as no relationship was found between treatment performance and recidivism. Quinsey, Khanna, and malcolm (1998) reported similar results, however again it is highlighted that this study was retrospective and adds to the earlier suggestions in regards to the importance of prospective data collections.

Despite these findings, research in to links between gains made in treatment and rates of reoffending has continued to grow and more promising results have been found. Beech et al. (2012) examined psychometric data for 413 child molesters who completed community based treatment in the UK. They categorised completers as responders or non-responders, according to their post-treatment psychometric profile. Offenders whose post-treatment scores were in the non-offender range on all three offence related measures and a minimum of three out of five Socioaffective measures were deemed to have 'responded' to treatment (n=135). Recidivism data indicated that 12% of the total sample had reoffended (general offences) within the 2 to 4 year follow up period. Twelve (9%) of the 135 'treated' profile offenders recidivated sexually and whilst the recidivism rates between the treated and untreated groups was not found to be statistically significant, a 40% reduction in recidivism was found for those deemed to have responded to treatment. Hence only tentative support can be drawn for successful treatment gains and reduced recidivism. Earlier results found by Scalora and Garbin (2003), also support this notion. They analysed data from 194 convicted child sex offenders some of whom received intensive cognitive behavioural treatment (n = 76). With an average follow up period of 54 months, 24.7% of offenders were found to have

been charged with a sexual offence post-release. For individuals who attended treatment, progress in treatment was assessed from statements within discharge documentation completed by variety of health professionals with individuals classified as successfully or unsuccessfully completing treatment. Successful completion was found to be related to reduced recidivism. Of note here is the reliance on clinical judgement for decision making regarding treatment progress. Whilst the authors note caution for a number of reasons regarding their results, the subjective nature of the assessment of progress appears likely to have the potential to influence results and adds weight to the argument for a robust method of measuring change.

Beggs and Grace (2011) employed three methods for measuring change in treatment to a sample of 218 adult males sexual offenders against children who had undergone institutionalized cognitive behavioural treatment. The three methods comprised change across a battery of psychometric variables measured pre and post treatment, post treatment ratings on the Standard Goal Attainment Scaling for Sex Offenders (Hogue, 1994) and change as measured by the Violence Risk Scale: Sexual Offender Version (Olver et al., 2007a). All measures of change were found to significantly predict recidivism and conclusions were therefore drawn regarding the effectiveness of treatment programmes which target dynamic risk factors. This paper is of particular importance to the current study and is therefore described more in-depth at a later point in the review, however this serves here to highlight the importance of the methodology of measuring change in accurately assessing the relevant of treatment gain. Other evaluations of the VRS:SO found similarly positive results in regards to relationships between change scores and recidivism (Olver, Beggs Christofferson, Grace, & Wong, 2014; Olver et al., 2007a) and add weight to the argument for including change information in post-treatment appraisals of risk.

Measurements of change

Overall, research on treatment change and recidivism have been inconsistent, however some encouraging results appear to be emerging. As often with research, early indicators of positive links between treatment gains and recidivism also bring with them unanswered questions and new directions. Common ground is found in these studies in regards to the need to ensure the robustness of the method used to assess treatment change. It is apparent that a concept central to much of the research regarding change in treatment is the notion of clinically significant and reliable change. Given the relevance to the current study, this concept will be reviewed in detail here with examples of its employment in sex offender research then defined.

Clinically Significant and Reliable Change

It is fair to assume that individuals who seek treatment in whatever form, are expecting the therapy to work and that they will experience some palpable change. From the 1970's a growing trend of interest emerged in to the effectiveness of treatments being offered across a variety of mental health and counselling fields. Both individuals and clinicians were becoming more and more interested in the outcomes of treatment interventions for individuals and there was a notable lack of models available for measuring this. Pre and post treatment group scores provide no information around how an individual fared in therapy, similarly statistically significance tests offer little in regards to clinical interest. Bain and Dollaghan (1991) commented on the benefits of this trend for clients, in regards to the increase in attention given to the quality of treatment, and the support this offers clinicians in their decision making regarding treatment planning. Further concerns adding to this trend were issues of a lack of common language in regards to change; as a focus on change between groups as opposed to individual change (Evans, Margison, & Barkham, 1998) and

focus appearing to be solely on statistical significance tests which bore little relevance for the practical importance of the effect (Jacobson, Follette, & Revenstorf, 1984).

The term 'clinical significance' emerged as a measurement criterion for psychotherapy. Definition and use of this term was varied with sole consensus appearing to be that clinical significance was different from statistical significance. Bain and Dollaghan (1991) quote clinical significance as being "clinicians' subjective judgements of the importance of the changes observed in a client undergoing treatment" (p. 264). They further comment that three dimensions contribute to such judgements; change resulting from treatment rather than other factors, change which can be shown to be real rather than random and change that is deemed important. Jacobson, Follette, and Revenstorf (1984) list a number of studies which provide further differing definitions of clinical significance including; "a large proportion of the clients improving" (Hugdahl & Ost, 1981), "an elimination of the presenting problem" (Kazdin & Wilson, 1978) and "a change which is recognizable to peers and significant others" (Kazdin, 1977) (p. 338). All of the above appear to be sensible requirements in terms of demonstrating a positive effect of specifically targeted treatment. However, the discrepancy between these and the many other definitions was vast and the need for a sound and user friendly consensus in the approach to treatment change measurement was therefore alarmingly apparent. Jacobson, Follette, and Revenstorf (1984) therefore acknowledged the challenges in regards to providing a standardized measurement of change which would be applicable to all clinical problems; however they believed that they could propose guidelines which could be useful across treatment modalities. They highlighted the importance of the amount of change being important as opposed to simply observing whether or not change had occurred. Their initial proposal involved the notion of change being deemed clinically significant if an individual's score at post treatment, fell within the range of the functional population. Three ways of determining whether a score met

the criteria for clinically significant change were proposed (Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991):

a) If the post test scores falls 2 standard deviations or more away from the dysfunctional mean (in the direction of functionality)

b) If the post test score falls 2 standard deviations or less below the functional population mean

c) If the post test score is statistically more likely to be drawn from the functional than the non- functional distribution

As norms are required to calculate (c), (a) and (b) are therefore the only choice when norms are not available. Jacobson and Truax (1991) comment that (a) is the more stringent choice the more overlap there is between the functional and dysfunctional distributions as it provides a more conservative measure. However, they highlight that choosing between (a) and (b) is arbitrary and therefore (c) is considered to be the best choice when distributions overlap (and norms are available) as it is not arbitrary. Of note, Jacobson, Follette, and Revenstorf (1986) proposed using confidence intervals around the cut-off point to allow for measurement error and avoid misclassification of clients, however Hansen and Lambert (1996) argue against this as many patients whose score pre-treatment falls within the confidence interval will not be able to be classified. Instead they argue that crossing the cut off is sufficient enough.

$$\text{Cut-off C} = \frac{s_0\ \bar{X}_1 + s_1\ \bar{X}_0}{s_0 + s_1} \tag{1}$$

Where $S_0 = Mean_{clin,}$ $X_1 = SD_{norm,}$ $S_1 = Mean_{norm,}$ $X_2 = SD_{clin}$

The proposed calculations of clinically significant change are clearly dependent on the accuracy of the measures. They also do not account for the amount of change that has

occurred. When distributions do not overlap, one could assume that change within treatment would be apparent for any individual who is considered to be in the normative range of functioning post treatment. However, often distributions do overlap or an individual may fall in to the high end of the dysfunctional distribution at the start of treatment, therefore true change within treatment is less conclusive. Jacobson, Follette, and Revenstorf (1984) therefore considered the issue of needing to assess whether the amount of change made by an individual was reliable and therefore not due to chance or an imprecise measure. They therefore proposed the Reliable Change (RC) index as a way to address this and accurately account for change that exceeded the margin of measurement of error. They proposed that an RC >1.96 would indicate that it was unlikely that the observed change was due to measurement error.

$$\mathbf{RC} = \frac{X^2 - X^1}{S_E}$$
(2)

X^2 = post test score, X^1 = pre-test score

S$_E$ (Standard error of measurement) describes the spread of the distribution of repeated performances that would be expected given that no actual change had occurred

Therefore Jacobson, Follette, and Revenstorf (1984) concluded that in order for a treatment effect to be considered reliable and significant, it must pass two criteria 1) the change must be proven to be statistically reliable (reliable change) and 2) the individual must pass from the dysfunctional to the functional population (clinically significant change). The authors were clear that their calculations were proposals and were not intended to be recommended as the definitive method of calculating change. Instead they insisted that their notion was more to ensure the opening of a dialogue with a view to the adoption of standard conventions within the field. Their work had precisely this effect. Christensen and Mendoza

(1986) highlighted concerns with the initial RC calculations in that "the standard error of measurement is an index of the dispersion of an obtained score about a true score" (p. 305). They argued that the above calculations are instead based on two obtained scores (pre and post) and do not take in to account the subject's true score pre-treatment. Christensen and Mendoza (1986) acknowledged that rarely in research situations are true scores acquired and therefore they sought to adapt the initial RC proposed calculations to take account of this. They instead proposed that the standard error of difference be used in the calculation of the difference between pre and post test scores to provide a calculation of Significant change (SC). They report this approach to be more stringent as it accounts for the fact that both the pre and post test scores contain measurement of error and therefore more accurately demonstrates whether or not an individual's change from pre to post treatment reflects true change. The standard error of difference is therefore the amount of difference you would expect to occur between two scores as a function of measurement error alone.

$$SC = \frac{X^2 - X^1}{S_{diff}} \qquad (3)$$

SC = significant change, X_1 = pre-test score, X_2 = post test score,

S_{diff} = standard error of difference between two test scores

Jacobson et al. (1986) responded, welcoming the above recommendations and agreeing to adopt the above formula in their future calculations. They further categorized individual's as 'recovered' (passed both cut off and RCI criteria), 'improved' (passed RCI but not cut off), unchanged (passed neither criteria) and 'deteriorated' (passed RCI criteria in a worsening direction).

Jacobson, Follette, Revenstorf, et al. (1984) and Jacobson and Truax (1991) both comment on limitations of their Clinical Significance formula and Reliable change index.

They highlight that the calculations are only as good as the outcome measures used and that frequently norms are not available for measures therefore restricting availability of standardized cut offs. They also comment that, for some conditions, joining the functional population may not be achievable or appropriate (e.g. schizophrenia) but that the change achieved may still make a significant difference to a client's life. Similarly Kazdin (2001) warned that such criteria may not take in to account the extent to which the changes made by an individual had a significant or important effect on their day to day life. He described scenarios in which change which would not be classed as clinically significant using the above criteria, may still have a profound and palpable impact on quality of life. Wise (2004) advised caution in regards to the criteria of returning to normal functioning. They comment that this may be appropriate for individuals with transient disorders but less relevant or achievable for individuals with chronic psychological symptoms. Jacobson, Follette, and Revenstorf (1984) also highlight that their criteria are realistic but conservative and quote a number of examples of studies in which the application of CSC demonstrated less favourable and more modest results in terms of the efficacy of treatment than standard inferential statistics (e.g. Jacobson, Follette, Revenstorf, et al., 1984; Jacobson, Wilson, & Tupper, 1988). Further conclusions around limitations of this approach include that the formulas proposed assume normal distributions and that examination of individual change scores does not allow for causal inferences that change is due to treatment alone. The authors therefore recommended that between group comparisons should also be used to highlight variability within the context of proportions of improved and deteriorated clients. They therefore concluded on the benefit of using such criteria as a complementary approach to the analysis of change alongside more traditional statistical methods.

Empirical evaluations of different proposed methods for computing change scores have been conducted and despite the above mentioned limitations, results have often favoured the

Jacobson and Truax (1991) approach. Speer and Greenbaum (1995) commented that this approach used straightforward calculations, avoided problems associated with residualised true scores and was well documented in research. Similar endorsements for this method were made by McGlinchey, Atkins, and Jacobson (2002) and Bauer, Lambert, and Nielsen (2004) with recommendations made specifically in regards to its utility in forensic populations (e.g. Friendship, Falshaw, & Beech, 2003) . This is likely to account for the apparent favoured use within sex offender research. Such calculations have been used for various purposes within sex offender research, including explorations of efficacy of treatment, outcome studies and recidivism. Table 1 summarizes the employment of clinically significant and reliable change methodology and other change methodology within forensic populations and specifically within sex offender research. Table 1 highlights the variety of ways in which calculations of change can be made even within one specific concept. As noted above, there are varying methods of calculating clinically significant change and even within that, some authors adopt their own interpretation cut off points (e.g. Nunes, Pettersen, Hermann, Looman, & Spape, 2014; Wakeling et al., 2013). This review indicates that utilizing the method of calculating cut off C appears to be the most widely used methodology. This is unsurprising given that Jacobson and Truax (1991) note this to be the most stringent method and therefore is likely to produce the most defensible results. However as noted in a number of studies, this relies on normative values being available for all measures which can therefore make it an unachievable methodology for some studies. Decision making regarding the calculation of reliable change was also found to vary across studies. Despite Jacobson and Truax (1991) acknowledging the usefulness of the Christensen and Mendoza (1986) proposed alteration to their initial reliable change index calculations, this method is not always adopted by studies seeking to calculate reliable change and clarity around the basis for this decision making is not always provided (e.g. Nunes et al., 2014). Whilst conclusions from such studies vary

greatly, overall there appears to be a trend of positive results in regards to the usefulness of this methodology in sex offender research with further research in to the links between clinically significant and reliable change and recidivism often reported as an important direction for future research. Given the mixed results to date, an important consideration would be the extent to which different methods of calculating change are more robust, generous or stringent than others and the implications this has for future assessments of change.

Table 1
A review of studies incorporating change methodology with sexual offenders

Study	Key research question	Change methodology used	Results
<u>Sex offender populations</u>			
Barnett, Wakeling, Mandeville-Norden, and Rakestrow (2011)	To examine the relationship between psychometric test change over treatment and sexual or violent reconviction	Post-treatment scores were deemed to be clinically significant if they were above or below the cut off (based on non-offender norms) in the direction of functionality. Reliable change index calculated using the Christensen and Mendoza (1986) method. Ten change categories were then used to classify individuals based on scores and reliable change status post-treatment	Main conclusion drawn was that treatment change was not associated with reduced sexual or violent recidivism
Beech, Beckett, and Fisher (1998)	To explore the longevity of treatment change and examine the relationship between treatment change and recidivism	Cut off C calculated using non-offender norms to determine CSC. RCI calculated using Jacobson's original formula utilizing standard error of measurement	Short and long term interventions were successful at producing change on nearly all measures
Beech and Ford (2006)	To examine the relationship between static and dynamic risk and reconviction	Cut off C calculated to determine CSC as per Jacobson and Truax (1991) methodology. Reliable change calculated using Christensen and Mendoza (1986) formula	None of the individuals deemed to have responded to treatment (i.e. those who achieved CSC) were reconvicted of a sexual offence

Study	Key research question	Change methodology used	Results
Beech and Hamilton-Giachritsis (2005)	To examine the relationship between therapeutic climate and effectiveness of CBT treatment for sexual offenders	CSC achieved if post treatment score crosses the cut off. Reliable change was not calculated with authors justifying this in regard to some offenders not having the level of deficits to achieve the necessary level of change required for RC.	No relationship found between treatment length and amount of treatment change. Significant results were found between treatment outcome and group cohesiveness and expressiveness
Beech, Mandeville-Norden, and Goodwill (2010)	To assess the short term effectiveness of treatment and the long term implications for sexual recidivism	Cut off C calculated according to Jacobson and Truax (1991) method. Reliable change calculation were made based on Christensen and Mendoza (1986) formula	Some support for hypothesis around lower recidivism rates for individuals deemed to have responded to treatment. A 40% reduction in recidivism was found for those who responded to treatment. Effect sizes were small but deemed nontrivial
Keeling, Rose, and Beech (2006)	To investigate the effectiveness of a custody based treatment programme for special needs sex offenders	RC calculated using Jacobson and Truax (1991) formula. CSC cut off score calculated for one measure using formula for overlapping populations (Jacobson and Truax 1991). Three outcome categories identified (Jacobson, Roberts, Berns, & McGlinchey, 1999). CSC not calculated for other measures due to a lack of norms and high standard deviations	Clinically significant reliable change observed in measures of offence supportive attitudes and victim empathy
Mandeville-Norden, Beech, and Hayes (2008)	To assess the effectiveness of community based therapeutic intervention for sex offenders	CSC calculated using Cut offs according to Jacobson and Truax (1991). Reliable change calculated according to Christensen and Mendoza (1986)	Significant proportion of sample demonstrated CSC post-treatment, most evident in Low need group. Conclusions drawn around efficacy of programme based on these results

Study	Key research question	Change methodology used	Results
Nunes, Babchishin, and Cortoni (2011)	To assess treatment change at group and individual level in treated Sex offenders using CSC and RC and to compare and contrast the two	Jacobson et al 1984 cut off c formula used to calculate CSC. Jacobson and Truax (1991) formula to calculate RC	Overall, more modest gains were noted for individual level than group level change analysis. However authors promote the use of CSC and RC calculations as a useful method of measuring treatment change at an individual level
Nunes et al. (2014)	To examine whether change on the Molest and Rape scales predicted recidivism in adult male sex offenders	Used formula for cut off B due to a lack of norms but modified it to 1SD above the functional mean as they believed 2SD would be too lenient for their sample. Only included individuals with dysfunctional pre-treatment scores. Used Jacobson and Truax (1991) formula for calculating RCI	Majority of sample had functional scores pre-treatment. Significant gains noted for majority of individuals with dysfunctional pre-treatment scores. Some evidence of recidivists' scores improving slightly more than non-recidivists
Olver, Beggs Christofferson, and Wong (2015)	To examine the use of CSC and RC with the VRS:SO and its implications for risk communication	Cut off calculated to distinguish the functional from the dysfunctional population. This was calculated as the upper threshold for scores on the VRS:SO for which there were no items identified as criminogenic (score of 17 and lower deemed to constitute the functional group – approximately 1SD below the mean). RC computed using standard error of difference (the more stringent Christensen and Mendoza formula). Four categories identified for CSC: Already okay, recovered, improved, unchanged	Already okay, recovered and improved group are essentially low, medium and high risk groups respectively. Significantly lower rates of recidivism found for already okay and recovered groups but the findings were concluded to be due to pre-treatment risk level. After controlling for risk, CSC for the improved category was associated with reduced recidivism

Study	Key research question	Change methodology used	Results
Wakeling et al. (2013)	To examine the relationship between treatment change, as measured psychometrically, and recidivism	CSC deemed to be achieved if post-treatment scores fell 1SD or more away from the pre-treatment mean in the direction of functionality. Original Jacobson and Truax (1991) method employed for calculating RCI. 5 change categories used to categorize individuals post-treatment: deteriorated, unchanged, improved, recovered, already okay	Conclusions regarding limited value in using CSC and RCI methodology in measuring progress in treatment. Significant association with recidivism found for some of the treatment change categories with lowest recidivism rates for *already okay* category and highest rates for *improved* category
General Forensic Populations			
Bowen et al. (2008)	To examine the relationship between clinically significant change and recidivism in a UK domestic violence sample	Jacobson et al (1984) cut offs for CSC Reliable change (Jacobson., 1999) Categories: normal, recovered, improved, deteriorated, regressed, unreliable	Clinically significant and reliable change was achieved by proportion of offenders. No association was found between clinically significant change and reoffending
Draycott, Kirkpatrick, and Askari (2011)	To assess effectiveness of treatment for dangerous and severe personality disorder using patient changes in treatment	CSC score is defined as one that exceeds the RCI on a given measure by at least one standard deviation. RCI calculations used are the original simpler version proposed by Jacobson and Truax (1991). Five point scale then used to classify results: significant improvement, improvement, no clear change, deterioration, significant deterioration	Largest group post treatment were those showing no change on any measure. Trend found towards improved scores on one measure. Conclusions drawn that these methods of assessing change are simplistic and useful and should be incorporated in to all psychiatric treatment
Schewe and O Donohue (1996)	To evaluate effectiveness of rape prevention interventions	Jacobson and Truax (1991) cut off calculations used to determine CSC	Clinically significant change was observed on three out of five measures

Study	Key research question	Change methodology used	Results
Tapp, Fellowes, Wallis, Blud, and Moore (2009)	To evaluate the impact of a cognitive skills programme (ETS)	Jacobson and Truax (1991) cut off calculations used to determine CSC Reliable change calculated according to Jacobson and Truax (1991)	Clinically significant change was achieved to the greatest degree for scales measuring problems/symptoms and social/life functioning. The least amount of reliable change was observed for these scales
Other change methodology			
Beggs and Grace (2011)	To determine whether treatment change would predict recidivism beyond pre-treatment assessments of risk	Standardized residual change scores were calculated by regressing the raw change scores on to the pre-treatment scores. Variance was then removed by calculating residuals from the regressions and standardizing the residuals for each variable	Positive results found for relationship between treatment gain and recidivism after controlling for static risk

Key relevant studies

As outlined thus far, a number of studies have investigated the relationship between treatment outcome and recidivism and specific change methodologies have played a key role in these. Of specific importance to this field of research is the development of the Violence Risk Scale – Sexual Offender Version (VRS:SO) (M. Olver, S. C. P. Wong, T. Nicholaichuk, & A. Gordon, 2007b). This instrument combines both static and dynamic risk variables to produce an assessment of risk of sexual recidivism and also identifies treatment targets and assesses treatment change and has been found to be predictive of reduced recidivism (Beggs & Grace, 2011). Data from the current sample has been used in multiple studies evaluating the effectiveness of the VRS:SO, with a key theme regarding assessing the usefulness of methods of accurately identifying and interpreting treatment change (e.g. Beggs & Grace, 2011; Olver, Beggs Christofferson, et al., 2014). Olver, Beggs Christofferson, et al. (2014) concluded that

changes in dynamic risk resulting from treatment completion, as measured by the VRS:SO, were related to reduced sexual and violent recidivism. In a further study of a sample of Canadian treatment completers, psychometric evaluations of treatment change were conducted (Olver, Kingston, Nicholaichuk, & Wong, 2014). Weak or nonsignificant relationships with recidivism outcomes were found for all measures but one (measures of hostility and aggression). However, further analyses incorporating residual change scores indicated that improvements in predictive validity were evident once pre-treatment scores were partialled out. Following these results, Olver, Kingston, et al. (2014) also conclude the usefulness of future research incorporating the reliable change index in to studies within this field alongside the usefulness of such analyses in regards to treatment programme efficacy.

A further evaluation of the VRS:SO has significant relevance to the current study as they incorporate the concept of clinically significant and reliable change (Olver et al., 2015). In their review of the literature to date, Olver et al. (2015) consider the various applications of the clinically significant change concept within sex offender research and conclude that, whilst initial value appears to be seen in regards to standardizing and interpreting change data from multiple scales, "its relevance to outcomes has yet to be convincingly established" (p.95).

Three other particular studies are key to this research and will therefore be reviewed in detail here. Allan, Grace, Rutherford, and Hudson (2007) studied a sample of adult male offenders (N= 495) who completed a prison based cognitive behavioral treatment programme in New Zealand. Specific aims of the study were to establish a methodology for the assessment of dynamic risk factors using an individual differences approach and to examine the relationship between dynamic risk factors, and recidivism. Their battery of psychometrics incorporated the following measures;

- The Abel-Becker Cognitions Scale (ABCS), (Abel et al., 1989)

- The Hostility Towards Women scale (HTW) (Check, 1984)

- The Rape myth acceptance (RMAS) (Burt, 1980)

- The Wilson Sex Fantasy Questionnaire (WSF) (Wilson, 1978)

- The Beck depression inventory (BDI) (Beck & Ward, 1961)

- The State-trait anxiety inventory (STAI) (Spielberger, 1983)

- The State Trait Anger Expression Inventory (STAXI) (Spielberger, 1988)

- The Social Self-esteem Inventory (SSEI) (Lawson et al., 1979)

- The Assertion Inventory (AI) (Gambrill & Richey, 1975)

- The Revised UCLA Loneliness scale (UCLS) (Russell, Peplau, & Cutrona, 1980)

- The Fear of Intimacy Scale (FIS) (Descutner & Thelen, 1991)

- The Adult Nowicki-Strickland Internal-External Control Scale (ANSIE) (Nowicki, 1983)

Exploratory factor analyses were conducted on psychometrics administered pre-treatment. Only a subset of the overall sample was included ($N = 232$) as psychometric data for all the measures was not available for the full sample. Four factors were identified – social inadequacy, sexual interests, anger/hostility and pro-offending attitudes – which summarized the overall sample. In order to examine links with recidivism, factor scores for each individual were computed using average standardized scores and overall deviance was calculated to measure the cumulative effect of these factors.

Results showed that all four factors were correlated with recidivism and that factors scores and overall deviance scores provided incremental predictive validity for sexual recidivism beyond the Static-99. Hence conclusions were drawn that individual scores from self-report psychometric measures can provide valid measures of dynamic risk with strongest links with recidivism found for Sexual interests and Pro-offending attitudes domains. Whilst this study does not examine treatment change or utilize clinically significant change

methodology, the current study utilizes the Allan et al. (2007) sample and four factor model and is therefore pertinent to this review.

In regards to treatment change, in a review, Beggs (2010) concluded that ratings of treatment outcome do have the potential to contribute to risk assessments of treated sex offenders, although cautioned that the variability of results across prior studies underscored the importance of ensuring the validity of measures of treatment outcome prior to applying them clinically. Beggs and Grace (2011) therefore applied three different methods of assessing gains made in treatment to a sample of child sex offenders who completed the Kia Marama programme, using a subset of the sample studied by Allan et al. (2007). These methods were: change on the four-factor psychometric risk assessment framework (developed by Allan et al., 2007); change on the three-factor Dynamic scale of the Violence Risk Scale: Sexual Offender Version (VRS:SO) (Olver et al., 2007a); and post treatment ratings of the attainment of treatment goals using a modified version of the Standard Goal Attainment Scaling for sex offenders (SGAS) (developed by Hogue, 1994). When difference scores on pre and post treatment psychometrics were analysed, Beggs and Grace (2011) found evidence of strong self-presentation or impression management bias at post treatment. Levels of deviance at pre-treatment were positively correlated with change scores, suggesting that offenders with more deviant pre-treatment scores had a greater opportunity to report post-treatment change. They therefore pursued an additional strategy, which involved the calculation of standardized residual change scores in which the pre-treatment scores were partialled out. Their results demonstrated that once the general tendency to report improvement was removed statistically, a valid measure of treatment change was obtained (Beggs & Grace, 2011). This contributed to their overall success in regards to demonstrating that specific gains made in treatment are significantly correlated with a reduction in risk for sexual recidivism.

In another recent study, Wakeling et al. (2013) also examined the relationship between treatment change and recidivism in sexual offenders as measured by psychometric self-reports. They identified a number of studies (e.g. Beech, Erikson, Friendship, & Ditchfield, 2001; Hedderman & Sugg, 1996) that had reported overall positive results in regards to positive change on post treatment psychometrics and reduced recidivism, including some studies which specifically used the concept of clinically significant change (CSC) to measure progress in treatment (e.g. Beech & Ford, 2006; Keeling et al., 2006). Thus Wakeling et al. (2013) investigated whether the CSC methodology could be useful for the assessment of treatment change with sexual offenders. They hypothesized that treatment change would be associated with reduced recidivism and that overall treatment change will add to the predictive validity of static risk measures for recidivism.

The sample studied by Wakeling et al. (2013) included 3773 sex offenders, all of whom had completed prison-based treatment programmes across England and Wales between 1996 and 2006 and who had been released from prison before April 2008. Treatment dropouts were not included in their sample. Risk of sexual offending was calculated for this sample using a modified version of Thornton's Risk Matrix 2000's indicator of sexual recidivism, RM2000-s (RM2000: (Thornton et al., 2003). They made these modifications because the intimate relationship and stranger items of this scale were not available for the entire sample. The Structured Assessment of Risk and Need framework (SARN, formerly SRA: Thornton, 2002), was also used to calculate dynamic risk for all participants as this is the predominant structured risk assessment tool used within the HM Prison Service in England and Wales. The psychometric battery employed in their study was administered pre and post treatment for all participants and comprised the following measures:

- The Multiphasic Sex Inventory (MSI; (Nichols & Molinder, 1984)
- The Entitlement to Sex scale (Hanson, Gizzarelli, & Scott, 1994)

- Sex with Children is Justifiable (Mann, Webster, Wakeling, & Marshall, 2007)

- The Women are Deceitful scale (NOMS) Rehabilitation Services Group, unpublished)

- The Locus of Control (Levenson, 1975)

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

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

- The Self-esteem scale (Webster, Mann, Thornton, & Wakeling, 2007)

- The UCLA loneliness scale (Russell et al., 1980)

- The Interpersonal Reactivity Index (Davis, 1980)

- The Impulsivity scale (Eysenck & Eysenck, 1978)

Wakeling et al. (2013) described the method of examining clinically significant change as seeking to establish whether or not an individual's score on a specific measure has shifted at post-treatment to be indistinguishable from that of an individual in the normal population (Jacobson & Truax, 1991). Wakeling et al. (2013) defined 'normal functioning' to be at least one standard deviation less (1 SD) from the pre-treatment mean. This is a deviation from the original cut off calculations recommended by Jacobson, Follette, and Revenstorf (1984) where two standard deviations were deemed necessary. Wakeling et al. (2013) reported choosing this particular cut off due to high standard deviations and low means for some of the measures. For individual offenders, CSC was assumed to have occurred if the psychometric score was 1 SD from the pre-treatment mean at post-treatment. They also calculated a reliability change index for each measure which indicated whether the extent of change shown by individual offenders was significant based on a standard error which depended on both the pre-treatment SD and the test-retest reliability of the measure. For each domain in the SARN (Thornton, 2002), offenders were then categorized as: deteriorated, unchanged, improved, recovered and already ok.

Their two year follow up period found reconviction rates of 1.7% for sexual offences, 4.4% for sexual and violent offences and a 12% overall general reconviction rate. They gave

the low base rates for sexual offences alone as a justification for their decision to use both sexual and violent reconviction data in their analysis. Overall their results demonstrated that, for most of the psychometric measures, the largest number of offenders fell in to the 'unchanged' group and the smallest number fell in to the 'deteriorated' category. They also found that only some measures showed a significant association between category and recidivism outcome and there were no significant difference in recidivism rates between any of the other treatment change groups, with the exception of the 'already ok' group.

As noted above, Wakeling et al. (2013) also used the SARN to measure dynamic risk. This tool identifies four clusters: Deviant sexual interest, pro-offending attitudes, socio affective problems and self- regulation problems and they used theoretical conclusions (not statistical analysis) to select which psychometric measures related to which of the four domains. Subsequently, participants who fell in to the 'recovered' or 'already ok' category for at least half of the measures in that domain were then further categorized in to 'change not required' group. Individuals, whose CSC and RCSC scores did not fall in to those categories, were considered as 'change still required'. Significant associations were found for three of the four domains (deviant sexual interests, socio affective problems and self-regulation problems) in regards to treatment change category and recidivism. Lower recidivism levels were identified in the 'change not required' (4.8%) groups compared to 'change still required' (8.1%).

Overall, Wakeling et al. (2013) concluded that their results provided only limited evidence to support the use of psychometric self-reports in the assessment of treatment change in sexual offenders using the CSC and RCSC methodology. However, they identify limitations of their work and considered their calculation of CSC using just 1SD as potentially being too stringent. Therefore concluding the utility of further research with a lower threshold of clinical

significance and overall potential benefits of examining different methods of calculating change to enable determinations around which is most appropriate.

Beggs and Grace (2011) results conflict with those of Wakeling et al. (2013) in regards to the utility of treatment change as measured through psychometric self-reports to contribute to the predictive validity of risk assessment. However, it is important to highlight that there were a significant number of differences in these studies. Most important of these is the difference in the methods used to measure treatment change – residual change scores (Beggs & Grace, 2011) and CSC/RCSC (Barnett et al., 2011; Wakeling et al., 2013). However, other significant differences are also evident, for example: the content and structure of the treatment programmes completed by participants, the sample size and makeup and the psychometric measures administered. Consequently, the reason for the different pattern of results across studies is uncertain at this point.

Rationale for the Current Research

The previous sections provide clear evidence of the variety of studies completed regarding psychometric change and recidivism. Methodological differences are apparent in each study, including sample size, treatment received, measures administered, follow up times and definitions of recidivism. Most apparent in regards to differences is the change methodology used to calculate treatment change and the methods used to define outcome. With such apparent conflicting results, there is a substantial need for additional research in this area. Therefore, in the present study we will apply the construct of clinically significant and reliable change to psychometric data for offenders who have completed the Kia Marama programme. Specifically, the general aim of this study is to compare different methods of CSC and RCSC calculations with the residual change methodology used by Beggs and Grace (2011). The rationale being that their results using RCZ found some significance where other studies have failed to do so using clinically significant and reliable change methodology. The

goal therefore is to determine whether CSC/RCSC also provides a useful measure of treatment change and provides significant results using a similar data set to Beggs and Grace (2011). If so, then the conflicting results of Beggs and Grace (2011) and Wakeling et al. (2013) may not be due to the method used to measure treatment change, but likely other factors. As noted above, these could include, among other factors, sample size, type/length/content of treatment programme delivered, offence type and follow up times. The decision to base comparisons on the Wakeling et al. (2013) study was based around the potential similarities of the population regarding all participants being incarcerated offenders and relative comparisons within the measures used. There was also limited variability within the treatment programmes completed. In the Wakeling et al. (2013) study all participants completed one of only two programmes, the present population all attended the same programme. Thus variability across treatment content was limited.

In this study, we therefore examined a sample of offenders who completed the treatment programme at Kia Marama. Clinically significant change and reliable change were calculated for all offenders. The sample size used was larger than that of the Beggs and Grace (2011) to provide for greater statistical power. The specific goals of this study include:

- To characterize offender progress overall on the administered psychometric battery in terms of two different calculations of Clinically Significant Change, a measurement of Reliable Clinically Significant Change , a measurement of a Reliable Change Index and Residual Change scores

- To compare five different methodologies for assessing change based on participants' pre and post treatment scores on the administered psychometric battery. The methods are outlined in detail below but overall comprise two methods of calculating clinically significant change, two methods of calculating reliable change and one calculation of residual change scores

- To determine which of the five identified methods of measuring change demonstrate the strongest correlation with recidivism

Chapter 2: Method

The Treatment Programme

The Kia Marama treatment unit was established in 1989 as the first prison-based treatment programme in New Zealand for sexual offenders. Kia Marama translates as "let there be light". The development of the programme was a direct result of attempts to respond to high rates of re-offending by child molesters, established by local research at around 25% by 1986 (McLean, 1990). The programme itself was devised by William Marshall with the original proposal based on Atascadero Sex Offender Treatment and Evaluation Programme (Marques, 1988). Programme content is based on relapse prevention framework and cognitive behavioural principles. This is one of only two specialised treatment units in New Zealand that are run by the Department of Corrections to offer therapeutic intervention to adult males who have engaged in sexually abusive behaviour. In order to be eligible for the programme, men must have a medium or minimum security classification and they must have either received a custodial sentence for a sexual offence against children under the age of sixteen or they must admit to such an offence. Admission to the programme is voluntary and selection excludes individuals with any active mental illness (although depression is common among programme participants) (Bakker & Westaway, 1998), or an IQ score lower than 70 (due to the cognitive nature of the content of the programme).

Participants undergo an initial two week assessment period involving psychometrics and clinical interviews. Focus is given to key areas such as; life management skills, interpersonal goals and ability to form satisfying intimate relationships, beliefs and attitudes about self, ability to regulate emotions (particularly the negative), capacity for empathy and perceiving victim harm, sense of responsibility for offences, attitudes to sex and what needs the individual believes are satisfied by his own deviant and non-deviant sexual activity

55

(Bakker & Westaway, 1998). A treatment plan is then formulated for each individual. Self-report psychometric measures are administered at this stage and post treatment.

The structured programme content is delivered in a group context via 9 hours of group therapy per week for 31 weeks with the sole facilitator led groups averaging eight members. Programme content is based around a cognitive behavioural therapy model including work focusing on identifying predisposing and precipitating factors which contributed to offending and relapse prevention planning. Individual therapy is minimal and described as "only enough individual therapy to allow a man to take part" (Bakker & Westaway, 1998). Non-therapy time is spent on assignments, therapy-related activities within the community or general prison work (e.g. kitchen and laundry). The treatment programme is made up of 8 key modules with individuals expected to complete all components: Norm building; Understanding your offending; Arousal reconditioning; Victim impact and empathy; Mood management; Relationship skills; Relapse prevention; Relapse planning and aftercare.

An evaluation by Bakker and Westaway (1998) concluded that treatment at Kia Marama had "a significant effect" (p. 2) and cited re conviction rates of 85 compared to a control group with rates of 21%. The authors comment that the results can be considered as particularly successful given that some of the control group were likely to have attended one to one counselling in prison. Bakker and Westaway (1998) also report significant changes on pre and post treatment psychometrics on measures of anger and sexual deviance and social skills, further concluding indications that treatment reduces cognitions and behaviours which contribute to offending against children. Of note, the control group time at large in the community was almost twice that of the Kia Marama graduates. The authors acknowledge therefore that the lower rates of recidivism could be a result of reduced time at large rather than treatment efficacy.

Within group analysis provided evidence of differences between treatment graduates who recidivated and those who did not. Analysis of variance results demonstrated that those who were reconvicted had significantly more previous convictions and prison sentences than those who did not. IQ scores for the reconvicted group were also commented on. The mean score for the non-reconvicted group was found to be close to the average for the normal population (100) whereas the mean score was found to be 8 points lower than the non-reconvicted group. The reconvicted group were also significantly more likely to report male victims or victims of both genders and they were almost twice as likely to report an onset of offending pre adulthood (age 20). Unsurprisingly, effects were also noted between pre and post treatment psychometric scores for the reconvicted and non-reconvicted groups. Reconvicted groups demonstrated an increase were more likely to report use of impersonal sexual fantasies and more likely to report attitudes supportive of offending (Bakker & Westaway, 1998). Overall the study provides evidence for the effectiveness of the Kia Marama treatment programme and also lends strong support to the importance of including analysis of within treatment change in outcome and efficacy studies. Beggs and Grace (2011) found similar evidence of efficacy of the Kia Marama treatment programme. Whilst programme efficacy was not a main target of their study, they demonstrated an association between treatment gain and recidivism and conclude that their findings provide indirect evidence of treatment efficacy. Similar findings were also indicated in an unpublished thesis by Moore (2012) who concludes that the Kia Marama treatment unit is successful in reducing the risk of recidivism for sexual offenders against children.

Sample

The participants of this study included adult males who completed the Kia Marama treatment programme. On entering the programme, all participants provided written consent for their file information to be used for research purposes.

Demographics

The population for this study comprised the sample used by Allan et al. (2007) of which the Beggs and Grace (2011) sample was a subset. Allan et al. (2007) original sample included clients who had entered treatment from the start of the Kia Marama unit in 1989 and were released prior to 1st February 2001. This totalled 557 offenders. Sixty two participants were omitted for various reasons, including 15 who were still incarcerated, 35 who failed to complete the programme, 2 who were deceased subsequent to release and 10 who were excluded due to a lack of information regarding their criminal history and/or demographic details. This gave a final sample of 495 and included the 242 men who were released before 1994 that were studied by Hudson, Wales, Bakker, and Ward (2002). Sample size differed across measures due to variance in the battery of psychometrics administered by Kia Marama over the period of time from which the data was collected.

The ages of participants ranged from 18 to 76 with an average age of 41 (SD = 12.2). 80.8% of the sample were of European descent, 16.7% NZ Maori and 2.5% other. The sample was approximately even in regards to offence type, with numbers comprising 52.3% incest offenders (for whom victims were exclusively from within their own family) and 47.7% extrafamilial offenders (who were unrelated to some or all of their victims).

Psychometric measures

Participants completed a battery of self-report questionnaires at both pre and post treatment. These measures assess across the four domains identified by Allan et al. (2007) factor analysis: F1 - Sexual interests, F2 - Social Inadequacy, F3 - Anger/Hostility, F4 - Pro-offending Attitudes. The battery comprised the following scales:

- The Abel-Becker Cognitions Scale (ABCS), is designed to measure distorted attitudes and beliefs about sexual offending against children (Abel et al., 1989). The scale

comprises 29 statements (all of which are consistent with pro-pedophile attitudes) which are rated on a 5 point scale for agreement. Overall scores range from 29 to 145 with scores being reversed, therefore a higher score indicates greater deviance. Tierney and McCabe (2001) reported an average ABCS (reverse scored) of 41.10 (SD = 11.20) for a non-offender community sample (N=40)

- The Hostility Towards Women scale (HTW) measures a cluster of negative beliefs about women including sex role stereotyping and acceptance of aggression against women (Check, 1984). The scale comprises 30 items scored as True or False, with higher scores indicating greater hostility towards women. Using a large, nationally representative sample of male college students (N= 2972), Malamuth, Sockloskie, Koss, and Tanaka (1991) provided normative data to show the average score on this scale was 7.29 (SD = 4.79). Hall (1989) found the average HTW score in a sample of 239 sex offenders was 6.82 (SD = 5.37)

- The Rape myth acceptance (RMAS) assesses beliefs supportive of sexual violence and aggression (Burt, 1980). Participants rate 19 beliefs about rape for agreement on a 7 point scale. Higher scores indicate greater support for myths about rape. Burt (1980) reported that the average score for a representative public sample (N=598) was 49.4 (SD=11.9)

- The Wilson Sex Fantasy Questionnaire (WSF) measures the frequency of a variety of sexual fantasies (Wilson, 1978). This scale consists of 40 items rated on a six point frequency scale. Subscales provide scores for four categories of sexual fantasy:

  Intimate themes (WSFIN) e.g. intercourse with a loved one

  Exploratory themes (WSFEX) e.g. group sex or partner swapping

Impersonal themes (WSFIM) e.g. sex with a stranger, voyeurism and

fetishism

Sado-masochistic themes (WSFSM) e.g. use of force or sexual humiliation

For each subscale, scores range from 0–50, therefore allowing the total score to

range from 0–200, with higher scores indicating more frequent fantasizing. Using a

sample of 116 male college students, Plaud and Bigwood (1997) reported the

following data: Intimate M = 31.7, SD = 9.3; Exploratory M= 14.3, SD = 7.9;

Impersonal M = 11.7, SD = 6.8; Sado-masochistic: M = 4.9, SD = 5.7; Total scores

M = 62.6 and SD = 23.9

- The Beck depression inventory (BDI), measures depressive symptoms using 21 items

  rated on a 4 point scale (Beck & Ward, 1961). Scores indicate depression levels as

  follows: <10 none or minimal depression; $10 - 18$ mild to moderate depression; $19 - 29$

  moderate to severe depression; $30 - 63$ severe depression. The BDI-II (Beck, Steer, &

  Brown, 1997) was used for participants who entered the program after 1997

- The State-trait anxiety inventory (STAI), (Spielberger, 1983) measures an individual's

  experience of anxiety currently (STAIS; state scale) and generally (STAIT; trait scale).

  Both scales consist of 20 items rated on a 4 point scale. Using a sample of working

  adult males, (Spielberger, 1983) reported norms for both scales: S scale M = 35.72 and

  SD = 10.40, T scale M = 34.89 and SD 9.19

- The State Trait Anger Expression Inventory (STAXI) is designed to measure several

  aspects of anger and anger expression (Spielberger, 1988). A total of 44 items are rated

  on a 4 point frequency scale, and form 5 major subscales. The authors describe the

  subscales as follows and report the following norms for each subscale using a sample of

  adult males:

60

State anger – (STAXS) intensity of anger at particular point in time (M = 11.29, SD = 3.17)

Trait anger (STAXT) – the degree to which an individual has a disposition to being angry (M= 18.65, SD = 4.81)

Anger expression (STAXE; anger out) – the degree to which anger is expressed towards people or objects in the environment (M = 14.41, SD = 3.33)

Anger suppression (STAXP; anger in) – the degree to which angry feelings are internalised (M = 15.36, SD = 3.92)

Anger control (STAXC) – the frequency with which an individual attempts to control anger expression (M = 26.20, SD = 4.26)

- The Social Self-esteem Inventory (SSEI) is designed to measure self-esteem in social situations as opposed to providing a global assessment of self-esteem in general (Lawson et al., 1979). The scale comprises 30 items rated on a 6 point scale with higher scores indicating higher self-esteem. The authors reported normative data, for an adult sample, of M = 312, SD = 21

- The Assertion Inventory (AI), is a scale designed to measure an individual's likelihood of making an assertive response as well as the degree of discomfort in a variety of specific situations (e.g. 'resist sales pressure') (Gambrill & Richey, 1975). The tool comprises 40 items rated separately on 5 point scales for discomfort (AI-D) and response probability (AI-RP). Higher scores indicate greater discomfort and lower response probability. The authors provided the following normative data for male undergraduate students: AI-D, M= 92.16, SD = 20.92; AI-RP, M= 104.22, SD = 15.95.

- The Revised UCLA Loneliness scale (UCLS) requires participants to make ratings on 20 statements regarding their perceived satisfaction or dissatisfaction with interpersonal relationships (Russell et al., 1980). The statements are rated on a 4 point scale with higher scores corresponding to greater loneliness. The authors reported the following normative data for a sample of male college students: M= 37.06, SD = 10.91

- The Fear of Intimacy Scale (FIS), measures individual's anxiety about close dating relationships (Descutner & Thelen, 1991). Participants are required to rate 35 items on a five point scale from 1 (not at all characteristic of me) to 5 (extremely characteristic of me). Higher scores indicate a greater fear of intimacy. (Doi & Thelen, 1993) provide normative data for a none university adult male sample: M=80.75, SD=23.70

- The Adult Nowicki-Strickland Internal-External Control Scale (ANSIE), is a measure of locus of control which defines locus of control as a generalised expectancy of control along an internal/external dimension (Nowicki, 1983). Participants answer 40 yes/no items with scores therefore ranging from 0 (internal locus of control) to 40 (external locus of control). (Nowicki, 1983) report that ANSIE scores have been found to be relatively free from social desirability bias and unrelated to gender or intelligence test scores with norms of M = 8.58, SD = 3.73 (Allan et al., 2007)

Static Risk

The Static-99 (Hanson & Thornton, 2000) was completed for all participants based on a review of file information. The Static-99 is one of the most widely used actuarial tools for assessing static risk in sex offenders. The tool consists of ten items, e.g. age, marital status, gender of victims and number of prior sex offences. It has been shown to have moderate to good predictive validity for sexual recidivism across a range of evaluation studies e.g. (Harris & Rice, 2003).

Calculations of Change

Clinically significant change and reliable change scores were calculated for individuals across all measures. CSC was calculated using the Jacobson and Truax (1991) method of establishing a cut off score based on normative data for each measure.

$$\text{Cut off} = \frac{(mean_{clin} \times SD_{norm}) + (mean_{norm} \times SD_{clin})}{SD_{norm} + SD_{clin}}$$

This provided four categories in to which individuals' pre and post treatment scores, respectively, could fall: a) above and below, b) below and below, c) above and above, d) below and above. Offenders who fell in to categories a) and b) were deemed to have achieved clinically significant change on the measure in question because their post treatment scores fell below the cut off. Hence they had demonstrated that, at post-treatment, they could be classified as scoring within the normal distribution of functioning. Offenders in category b) would be deemed to be non-deviant at pre-treatment (i.e. their pre-treatment score fell in to the range of the normal distribution) and therefore one would expect them to remain so following treatment. It could therefore be concluded that these individuals had not made any change in treatment due to the fact that no change was necessary, i.e. they were functional at pre-treatment. Category c) indicates that an individual has scored above the cut-off at both pre and post treatment and therefore would be deemed to have been deviant at pre-treatment and remained there despite completing treatment. Category d) defines individuals whose pre-treatment score fell in to the range of the normal distribution but who, at post-treatment, appeared to have deteriorated with their score falling above the cut off, in the deviant distribution.

It is category a) which defines offenders who were deviant pre-treatment and have made significant enough change for their post treatment score to fall in to the normal range. It can be said that this is ideally what treatment is hoping to achieve and that such shifts

demonstrate a positive treatment outcome and contribute towards efficacy of treatment. However, scores across repeated test administration can be due to measurement error and therefore the reliability of such change is crucial in order to draw more concrete conclusions about changes from deviant to non-deviant functioning. Reliable change was therefore calculated for scores across all measures, thus providing information as to whether the change observed was not due to chance and instead reflects genuine change. Christensen and Mendoza (1986) proposed a more stringent formula for the calculation of reliable change than their earlier colleagues (Jacobson, Follette, & Revenstorf, 1984) which incorporates a measurement of the standard error of difference. This approach was used to calculate reliable change across the sample. Scores exceeding 1.96 times the standard error of difference were deemed to be reliable i.e. unlikely to occur more than 5% of the time and therefore not due to measurement error alone. The standard error of difference was calculated using Evans et al. (1998) formula $SE_{diff} = SD_1\sqrt{2}\sqrt{1}\text{-}r$ (where SD_1 is the standard deviation of the baseline observations and r is the reliability of the measure). RCSC therefore denotes scores who achieved positive results in regards to both CSC and reliable change.

A third method of calculating change was employed. This mirrored the method used by Wakeling et al. (2013) to calculate clinically significant change (denoted in this study as WCSC). Wakeling et al. (2013) defined 'normal functioning' to be at least one standard deviation less (-1 SD) than the pre-treatment mean in the direction of functionality. The Wakeling et al. (2013) method of calculating reliable change, $RCI = X^1 - X^2/SE$ (denoted WRCI for this study) was also applied to scores for each measure indicating whether the extent of change shown by individual offenders was significant based on a standard error which depended on both the pre-treatment SD and the test-retest reliability of the measure. Of note, the above formula was produced in the published article as differing from the original formula cited. It has been assumed here that the absence of relevant parenthesis is a printing

error and thus an assumption has been made that their calculations were made using the correct formula and subsequently calculations in this study have been made using the correct version.

Lastly, Beggs and Grace (2011) method for calculating Residual Change Scores (denoted BRCZ here) was applied. This involved the calculation of standardised residual change scores in which the pre-treatment scores were partialled out. Raw change scores were regressed on to the pre-treatment scores. Residuals were calculated to remove variance (obtained change score – predicted change score) and then standardized for each variable.

Data Analysis

Firstly, this study will use descriptive statistics to analyse how the overall sample performs in terms of the Concepts of Clinically Significant Change, Reliable Change and Residual Change scores. Different ways to calculate clinically significant and reliable change will be applied, including using non offender normative samples (used by Allan et al., 2007) versus using the pre-treatment mean as the basis of the normative score (Wakeling et al., 2013). Correlations with sexual recidivism will then also be analysed for all five methods of calculating change for individual psychometric variables, Allan et al. (2007) dynamic risk factor domains and average overall change scores. Hierarchical logistic regression will then be used to identify if any one method of calculating change performs better in regards to predicting sexual recidivism. Overall, results should provide a clear picture of the relative utility of different methods for assessing treatment progress based on change scores and the power of these methods to provide incremental predictive validity regarding sexual recidivism.

Table 2

Table of psychometrics indicating which measures load on to each of the four factors identified by Allan et al. (2007)

Factor	Psychometric test
Social Inadequacy (F1)	Social Self-esteem Inventory
	Beck Depression inventory
	State-Trait anxiety inventory
	State
	Trait
	Social Self-esteem Inventory
	State-Trait Anger Expression Inventory
	Suppression
	Assertion Inventory
	Response Probability
	UCLA Loneliness scale
	Fear of Intimacy Scale
	Hostility towards women scale
Sexual Interests (F2)	Wilson Sex Fantasy Questionnaire
	Exploratory
	Intimate
	Impersonal
	Sadomasochistic
Anger/Hostility (F3)	State-Trait Anger Expression Inventory
	State
	Trait
	Expression
	Control
Pro-Offending Attitudes (F4)	Abel-Becker Cognitions Scale
	Rape Myth Acceptance Scale
	Adult Nowicki-Strickland Internal-External Control Scale

Recidivism

Reconviction rates were examined for the sample (N = 495). The average follow up

period was 13.18 years with a minimum of 7.55 and maximum of 17.96 years. During this

time, a total of 9.9% (*n* = 49) of offenders were reconvicted for a new sexual offence, 9.7% (*n*

= 48) for a new violent offence and 16.4% (*n* = 81) for a new general offence (nonsexual and

nonviolent). For those that reoffended sexually, the average time until a new offence was 2.55

years.

Psychometric Change

Table 3 shows mean pre-treatment and post treatment scores obtained for the sample

across all psychometric variables. All variables were scored such that higher values indicated

greater deviance or risk. Effect sizes for comparing pre vs. post treatment scores with paired-

sample *t* tests were calculated for all variables to provide an initial measurement of treatment

change, with medium to large effects sizes indicating change. Positive effect sizes indicate

pro-social change, that is, a less deviant or dysfunctional score post treatment. Results

demonstrated statistically significant effect sizes in the direction of pro-social change for all

measures with the exception of the Stait Trait Anger Expression Inventory control subscale

(STAXC) and Wilson Sex Fantasy Questionnaire, intimacy scale (WSFIN). The largest effect

size was obtained for ABCS (*d* = .84) which measures distorted cognitions related to children

and sexual behaviour. Other medium to large effect sizes were observed, which indicates that

positive change was achieved for many of the variables; hostility towards women (*d* = -.53),

rape myth acceptance (*d* = -.59, depression (*d* = -.77), state and trait anxiety (*d* = -.52, *d* = -

.54 respectively), anger suppression (*d* = -.56), assertiveness (*d* = -.52), loneliness (*d* = -.58)

and locus of control (*d* = -.54). The results indicate medium to large gains in treatment were

Table 3

Means, Standard Deviations, N and minimum and maximum score for Pre-Treatment and Post-Treatment Psychometric Variables, Normative Cut Off scores and Effect sizes (Cohen's d)

Test	Pre-treatment					Post-treatment					Normative Cut Off	Effect Sizes[1]
	Mean	SD	N	Min	Max	Mean	SD	N	Min	Max		
ABCS	122.27	15.96	471	52	145	135.21	10.39	441	29	97	45.48	0.84 ***
HTW	11.89	6.39	471	0	30	8.57	6.04	440	0	28	9.26	0.55***
RMAS	48.49	19.05	465	19	112	37.63	17.51	449	19	127	49.05	0.69***
WSFEX	10.89	8.49	471	0	44	7.80	7.11	438	0	36	12.66	0.43***
WSFIN	23.69	11.15	471	0	48	23.55	11.50	438	0	50	28.06	0.03
WSFIM	11.39	7.86	471	0	39	8.12	6.53	438	0	39	11.56	0.45***
WSFSM	4.12	5.72	471	0	37	2.89	4.98	438	0	32	4.51	0.22**
BDI	15.99	9.73	454	0	51	8.96	8.53	355	0	46	12.03	0.72***
STAIS	39.98	13.65	472	10	80	33.24	12.13	455	20	80	37.56	0.44***
STAIT	44.33	12.27	472	15	75	37.97	11.28	455	20	78	38.93	0.50***
STAXS	13.60	6.27	425	10	40	11.72	4.00	401	10	36	12.06	0.28***
STAXT	18.93	6.00	427	10	40	17.26	5.35	401	10	40	18.77	0.30***
STAXE	15.97	4.65	425	8	43	15.48	3.77	402	8	32	15.06	0.11*
STAXP	18.24	4.56	416	8	31	15.72	4.41	402	8	29	16.69	0.48***
STAXC	22.04	5.66	425	8	32	22.85	5.53	402	8	38	24.41	0.14
SSEI	112.96	27.94	474	37	178	120.81	26.35	450	52	180	123.84	0.29***
AIRP	113.29	21.97	417	43	182	101.04	24.83	382	10	200	108.03	0.50***
FIS	93.76	23.25	284	36	152	87.13	22.07	288	37	146	87.31	0.31***
UCLS	46.13	10.14	383	22	76	39.97	11.05	372	20	73	41.76	0.55***
NSIES	15.72	6.03	473	1	32	12.53	5.82	445	1	28	11.30	0.61***

achieved based on the offender self-reports, specifically relating to a decrease in pro-offending attitudes and enhancement of social adequacy.

Different measurements of change

A major goal of our study was to compare different measures of treatment change used with the same sample. Table 4 shows the percentage of individuals who demonstrated *CSC*, *RCSC*, *WCSC* and *WRCI* across all measures. Different *N* values were found across the various methods of calculating change. This indicates discrepancies in the sample sizes across the measures. This is likely to have occurred when no pre-treatment score was recorded but a valid post-treatment score was found. This was relatively unusual but did occur in some cases. Thus, for example, a valid CSC score could be returned as the post-treatment score could have fallen in to the normal range but reliable change could not be calculated owing to the absence of a pre-treatment score.

Over 50% of individuals achieved CSC on each measure with the exception of the NSIES however this was close to 50% (49.55%). Results for the ABCS (83.64%) indicate that individual scores on this measure produced the highest degree of CSC compared with other measures, with results on the STAXS also showing similarly high results (82.04%). Overall, 65.31% of the population achieved CSC. Percentages of offenders who achieved RCSC were overall much lower than CSC and ranged from 42.76% on the ABCS to 2.51% on the STAXC. On average across all measures 19.43% of the population achieved RCSC.

As described in chapter 2, Wakeling et al. (2013) used different methodologies to calculate clinically significant and reliable change, denoted here as WCSC and WRCI. Percentages across all variables are also presented in Table 4 for these methods of change. Over 70% of individuals achieved WCSC on all measures. The highest return was ABCS (100%) with BDI also scoring highly (95.21%) and the lowest being SSEI (74.22%). In total,
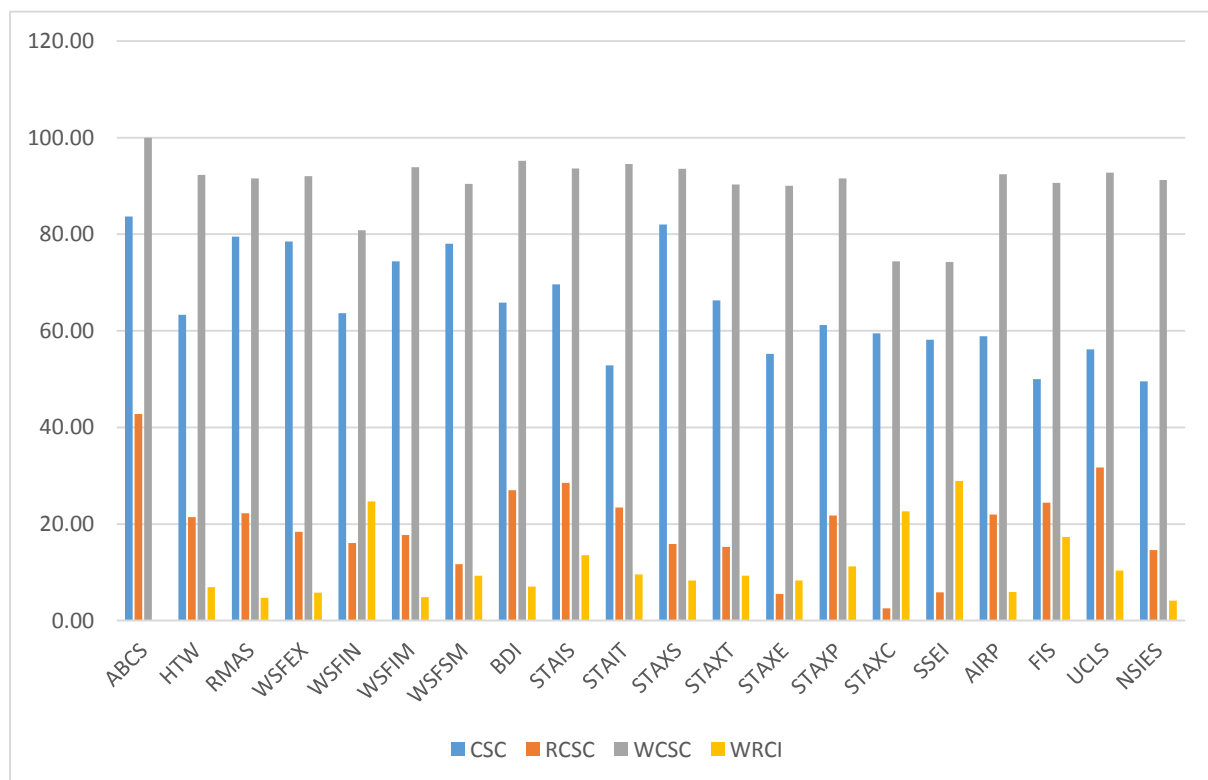
Table 4

Percentages of participants who demonstrated positive results in regards to different methods of change measurement

	CSC			RCSC			WCSC			WRCI		
Test	*N*	*n*	*%*	*N*	*n*	*%*	*N*	*n*	*%*	*N*	*n*	*%*
ABCS	440	368	83.64	435	186	42.76	441	441	100.00	436	0	0.00
HTW	439	278	63.33	434	93	21.43	440	406	92.27	435	30	6.90
RMAS	448	356	79.46	441	98	22.22	449	411	91.54	442	21	4.75
WSFEX	437	343	78.49	430	79	18.37	438	403	92.01	430	25	5.81
WSFIN	437	278	63.62	429	69	16.08	438	354	80.82	430	106	24.65
WSFIM	437	325	74.37	429	76	17.72	438	411	93.84	430	21	4.88
WSFSM	437	341	78.03	429	50	11.66	438	396	90.41	430	40	9.30
BDI	354	233	65.82	341	92	26.98	355	338	95.21	341	24	7.04
STAIS	454	316	69.60	449	128	28.51	455	426	93.63	450	61	13.56
STAIT	454	240	52.86	449	105	23.39	455	430	94.51	450	43	9.56
STAXS	401	329	82.04	397	63	15.87	401	375	93.52	397	33	8.31
STAXT	401	266	66.33	399	61	15.29	401	362	90.27	399	37	9.27
STAXE	402	222	55.22	398	22	5.53	402	362	90.05	398	33	8.29
STAXP	402	246	61.19	391	85	21.74	402	368	91.54	391	44	11.25
STAXC	402	239	59.45	398	10	2.51	402	299	74.38	398	90	22.61
SSEI	449	261	58.13	445	26	5.84	450	334	74.22	446	129	28.92
AIRP	382	225	58.90	373	82	21.98	382	353	92.41	373	22	5.90
FIS	288	144	50.00	266	65	24.44	288	261	90.63	266	46	17.29
UCLS	372	209	56.18	328	104	31.71	372	345	92.74	328	34	10.37
NSIES	444	220	49.55	438	64	14.61	445	406	91.24	439	18	4.10
Overall Average	414.00	271.95	65.31	404.95	77.90	19.43	414.60	374.05	90.26	405	43	11.20

scores on 16 of the 20 variables demonstrated over 90% WCSC with an overall average of

90.26%. Results were demonstrably different in regards to reliable change (WRCI). Table 4

shows that 0% of individuals achieved WCI on the ABCS measure with the SSEI returning

the highest results (28.2%) and an overall average of just 11.2%.

As can be seen in Figure 1, overall, the two measurements of clinically significant

change yielded higher average scores across all measures than the two types of reliable

change measurement. Notable differences are also apparent between the two types of

clinically significant change calculations with WCSC demonstrating higher levels of

clinically significant change across all measures than CSC. In regards to reliable and

significant change, the less stringent of the two methods WRCI, demonstrated lower averages

of reliable change than RCSC on 15 of the 20 measures.

Figure 1
*Percentages of participants who demonstrated positive results across four methods of
measuring change*

Measurements of percentages of change across factors

Table 5 demonstrates the percentages of change achieved by four methods of calculating change across the dynamic risk factors identified by Allan et al. (2007). Overall averages were calculated for each of the four factors for all four calculation methods. For CSC, the lowest percentages were found for the Social Inadequacy (F1; 58.44%), followed by Anger/Hostility (F3; 65.73%) and then Pro-offending attitudes (F4; 67.06%). The highest overall average percentages of CSC were found for F2, Sexual Interests (73.46%). In contrast, for calculations of change using the WCSC method, lowest percentages were found for Anger/Hostility (F3; 87.06%) followed by Sexual Interests (F2; 89.27%) and then Pro-offending attitudes (F4; 90.36%) with highest averages found for Social Inadequacy (F1; 90.84%). As can be seen, for this method of calculating change, there was only a range of 3.78% between average percentages of clinically significant change across all four factors.
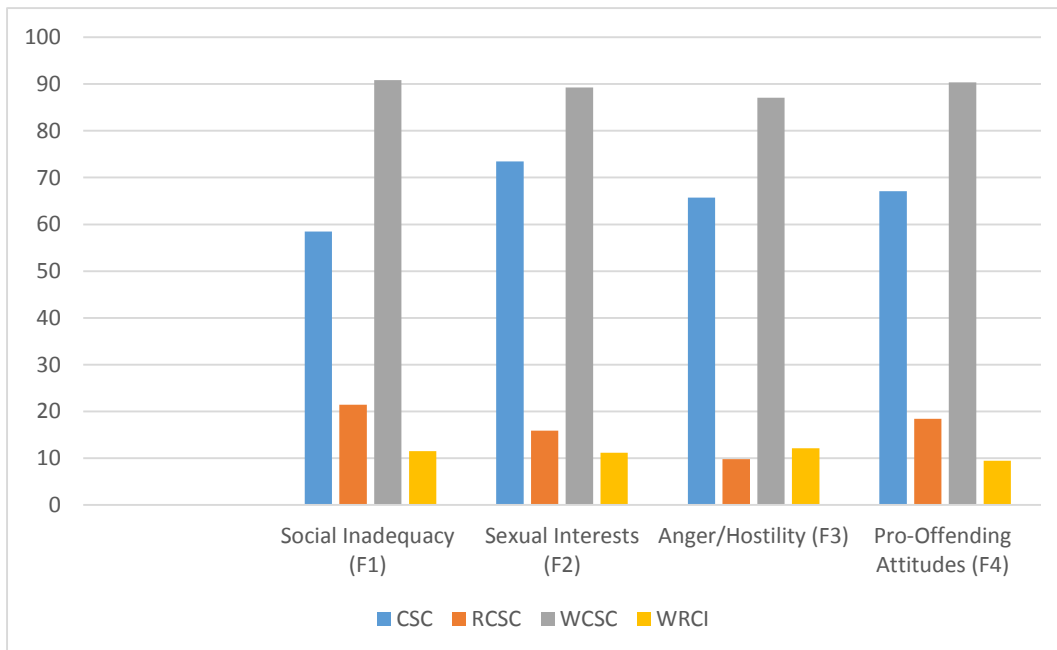
Table 5
Overall average percentages of change across four factors

Factor	Method of Calculating Change			
	CSC	RCSC	WCSC	WRCI
Social Inadequacy (F1)	58.44	21.43	90.84	11.49
Sexual Interests (F2)	73.46	15.89	89.27	11.16
Anger/Hostility (F3)	65.73	9.81	87.06	12.12
Pro-Offending Attitudes (F4)	67.06	18.39	90.36	9.43

For the reliable change methodology, again patterns of results across the four factors were inconsistent. For RCSC, lowest averages were found for Anger/Hostility (FE; 9.81%), followed by Sexual Interests (F2; 15.89%) and then Pro-offending attitudes (F4; 18.39%) with highest overall average found for Social Inadequacy (F1; 21.43%). For WRCI, lowest percentages were found for Pro-Offending Attitudes (F4; 9.43%), followed by Sexual Interests (F2; 11.16%) and then Social Inadequacy (F1; 11.49%). The highest overall average percentages for WRCI were found for F3, Anger/Hostility (12.12%).

There was no consistent pattern across all four methods of calculating change in regards to which factors demonstrated the most observed change. However for both RCSC and WCSC, results were consistent: Social Inadequacy (F1) demonstrated the highest levels of change (RCSC 21.43%, WCSC 90.84%) followed by Pro-Offending Attitudes (F4) (RCSC 18.39%, WCSC 90.36%), Sexual Interests (F2) (RCSC 15.89%, WCSC 89.27%) and finally Anger/Hostility (F3) (RCSC 9.81%, WCSC 87.06%). Figure 2 demonstrates that whilst there were obvious differences in average scores between the different methods of calculating change, patterns observed indicate limited range within each method. This suggests that for each method considered individually, the differences between scores on each factor were not significantly different.

Figure 2
Overall average percentages of change across four factors for four methods of measuring change

Relationship between treatment change calculation methods and sexual recidivism

Next we assessed the extent to which the different measures of treatment change were able to predict sexual recidivism. Table 6 presents the results of an analysis of correlations

Table 6

Correlations with recidivism for five methods of measuring change

Test	CSC	RCSC	WCSC	WRCI	BRCZ
ABCS	-.146**	0.03	.c	.c	-0.09
HTW	-.111*	0.06	-.103*	0.05	-0.04
RMAS	-0.08	0.01	-0.03	0.03	-0.01
WSFEX	-.143**	-0.07	-.188**	.096*	-0.04
WSFIN	-0.06	-0.08	-0.01	-0.04	-0.01
WSFIM	-.096*	-0.07	-.109*	0.00	-0.05
WSFSM	-.122*	0.03	-.129**	0.07	-.119*
BDI	-.127*	0.03	-.149**	0.10	-.127*
STAIS	-0.04	0.03	-.099*	0.02	-0.05
STAIT	-.117*	.137*	-0.04	0.02	-.133**
STAXS	-.158**	0.12	-0.07	0.07	-0.09
STAXT	-.167**	0.04	-.216**	0.08	-.145**
STAXE	-0.07	0.10	-.137**	0.04	-0.05
STAXP	-.150**	0.07	-.195**	.147**	-.120*
STAXC	0.04	-0.03	0.05	-0.02	-0.04
SSEI	-0.02	-0.07	0.00	0.06	0.06
AIRP	-0.08	0.00	-.156**	.108*	-.112*
FIS	-0.07	0.01	-.124*	-0.02	-.121*
UCLS	-.136**	0.02	-0.09	0.00	-0.05
NSIES	-.101*	-0.07	-.134**	-0.01	-.103*

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)
a. Cannot be computed because at least one of the variables is constant

exploring the relationship between change outcome on all psychometric variables and sexual

recidivism. Findings indicated that positive change on a number of psychometric variables was

positively correlated with sexual recidivism. Positive values in change scores indicated

change in a pro-social direction, that is, an individual was deemed less dysfunctional post-

treatment. Correlations with recidivism were therefore expected to be negative (i.e., smaller

pro-social change would be associated with higher rates of recidivism). Significant

correlations were found for a number of measures across the five methods of measuring change, however substantial differences were found in the pattern of results. Both measurements of clinically significant change yielded significant correlations for 12 of the 20 measures, however results were not significant for the same measures for each method of calculation. Consistency was found for only 8 of the measures with both CSC and WCSC producing significant correlations for measures of hostility towards women, three subscales of the Wilson Sex Fantasy questionnaire (exploratory, impersonal and sado-masochistic), measures of depression, two subscales of the State Trait Anger Expression Inventory (trait and anger suppression) and measures of locus of control.

In regards to reliable change, RCSC calculations were only significantly correlated with recidivism for one variable (STAIT) and WRCI was found to be significantly correlated for just 3 variables (WSFEX, STAXP and AIRP). Residual change scores (BRCZ) were significantly correlated with recidivism for 8 of the variables (WSFSM, BDI, STAIT, STAXT, STAXP, AIRP, FIS, NSIES). Four of the 20 measures demonstrated no positive relationships with sexual recidivism for any of the methods of calculating change (RMAS, WSFIN, STAXC, SSEI).

As described above, correlations with recidivism were expected to be negative, demonstrating that greater pro social change would result in a decrease in recidivism. Most of the significant correlations were indeed negative as expected. However, all significant correlations for both methods of calculating reliable change were found to be positive. This suggests that reliable change in a pro social direction is associated with an increase in recidivism.

As noted in Table 6, valid scores could not be computed for correlations between sexual recidivism and WCSC and WRCI scores for the variable ABCS. All scores for the ABCS

achieved clinically significant change as measured by the Wakeling methodology and no scores were deemed to have achieved reliable change (WRCI). These measures were thus deemed constant for the purposes of correlations and therefore no calculations could be made.

Correlations across factors

Correlations between average change scores and sexual recidivism across factors are demonstrated in Table 7. In regards to calculations of clinically significant change, positive correlations were found for all four factors for CSC and WCSC. However, results again differed for reliable change. No correlations were found for RCSC and only F1, social inadequacy returned a positive correlation with WRCI. In regards to residual change calculations, F1, Social inadequacy and F3, Anger/Hostility were positively correlated with sexual recidivism. As expected, correlations were negative, indicating a relationship between increased pro social change and reduced recidivism. Again there were exceptions to this in regards to the significant correlation found between reliable change and Social Inadequacy (F1). This was the only correlation found to be positive, again suggesting that change measured with this method is associated with an increase in recidivism.

Table 7
Correlations with sexual recidivism for average scores across factors for five methods of measuring change

| Psychometric Factor | Method of Change | | | | |
	CSC	RCSC	WCSC	WRCI	BRCZ
F1 Social Inadequacy	-.152**	0.004	-.199**	.134**	-.099*
F2 Sexual Interests	-.137**	0.08	-.149**	0.023	-0.07
F3 Anger/Hostility	-.159**	-0.034	-.151**	0.064	-.117*
F4 Pro-Offending Attitudes	-.118*	0.031	-.096*	0.002	-0.085

** Correlation is significant at the 0.01 level (2-tailed)

Overall change scores

Table 8 shows correlations between sexual recidivism and overall average change scores as measured by the five methods of calculating change. All correlations were significant

(at the 0.01 level) with the exception of RCSC. As expected, negative correlations were

achieved. The exception was again relating to calculations of reliable change, WRCI, where

again a positive correlation was found.

Table 8

Correlations with sexual recidivism for overall change scores across five methods of measuring change

| | Method of Change | | | | |
	CSC	RCSC	WCSC	WRCI	BRCZ
Sexual Recidivism	-.206**	0.025	-.228**	.127**	-.130**

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

The above analyses show that four of the five methods of calculating change were

correlated with sexual recidivism (CSC, WCSC, WRCI, BRCZ). Results thus far appear to

suggest that WCSC shows promise as the most effective predictor of sexual recidivism. A

key aim of this study was to identify, which, of the five methods of calculating change

demonstrate the most usefulness in predicting recidivism. To further address this question, a

series of hierarchical logistic regressions were conducted. Logistic regression analyses were

performed (incorporating each of the four methods of calculating change) with sexual

recidivism as the dependent variable. RCSC was not included in the analysis as overall

average change scores calculated by this method were not found to be significantly correlated

with recidivism.   For these analyses, WCSC was entered at the first step and CSC, WRCI, or

BRCZ was entered at the second step. The question was whether any of these latter three

measures would contribute significantly to predictive validity for recidivism beyond the

WCSC.

Results are shown in Table 9. The increase in predictive power was not significant for of

the Chi squares calculated in the regression model, indicating that clinically significant change

as calculated by the Wakeling methodology (WCSC) is the measure with the best predictive

validity for recidivism.

Table 9

Results of logistic regression analyses for sexual recidivism with overall average change scores
for four methods of calculating change based on self-report psychometrics

Variable	Chi-Square	Sig	B	e^B
WCSC		0.035	-0.031	0.969
CSC	3.31	0.069	-0.016	0.984
WCSC		0	-0.048	0.953
WRCI	0.039	0.844	0.003	1.003
WCSC		0.001	-0.051	0.951
BRCZ	0.005	0.945	0.022	1.022

(* p < .05, ** p < .01, *** p < .001)

Chapter 4: Discussion, Limitations & Conclusions

Summary of the study

The main purpose of this study was to compare the utility of five different methodologies for assessing change from psychometric self-reports and to determine which gave measures with the best predictive validity for recidivism. These methods included the Jacobson and Truax (1991) method of establishing a cut off score based on normative data for each measure; change defined as clinically significant when the post treatment score fell 1SD away from the pre-treatment mean in the direction of functionality; two methods of calculating reliable change: the Jacobson, Follette, Revenstorf, et al. (1984) calculation adopted by Wakeling et al. (2013) and the more stringent formula proposed by Christensen and Mendoza (1986); and finally residual change score calculations replicating the methodology adopted by Beggs and Grace (2011). Wakeling et al. (2013) did not find positive results for the utility of CSC as a useful method of change, therefore we applied their methodology for calculating clinically significant and reliable change (WCSC and WRCI) to the current data set to provide an independent test of their method.

Discussion of findings

Results showed that, in regards to overall change scores, four of the five methods of assessing change; CSC, WCSC, WRCI, BRCZ, were significantly correlated with recidivism, such that greater change was associated with a reduction in risk of reoffending. This is consistent with results from Wakeling et al. (2013) who found that overall treatment change ratings were associated with recidivism and Beggs and Grace (2011) who reported that measures of treatment change as measured psychometrically significantly predicted recidivism. In regards to identifying the most effect measure of change, logistic regression analyses showed that clinically significant change as calculated by the Wakeling

methodology (WCSC) was the measure with the best predictive validity for sexual recidivism.

The only method not found to correlate with sexual recidivism was RCSC. In order to achieve clinically significant change, scores must only satisfy one requirement, which is to fall in to the normal population range post treatment. However, establishing that change is genuine and not due to chance requires more stringent criteria. Therefore results were expected to reflect the notion that reliable change is more difficult to achieve, with clinically significant change (from either method of calculation) being achieved by the sample in abundance of levels of reliable clinically significant change. When comparing the two methods of calculating reliable change, this study employed the more stringent formula (RCSC) recommended by Christensen and Mendoza (1986), and adopted by N. Jacobson and Truax (1991), which takes in to account measurement error and seeks to establish results based on true scores. Wakeling et al. (2013) opted not to use this more robust method of measuring reliable change with the justification that they perceived their method to be stringent enough. As seen here, the less stringent calculation does achieve significant results in regards to correlation with sexual recidivism however this must be considered alongside non-significant results for the stricter formula. Given that the premise of reliable change is to conclude that observed change is not due to chance, it may be concluded that this is the most stringent, or accurate method of assessing genuine treatment change. Because our results indicated that reliable clinically significant change (as measured with the more stringent formula) was not correlated with recidivism, one conclusion is that overall genuine change on dynamic risk factors, as measured psychometrically, is not predictive of reduced recidivism. However, given that the Wakeling reliable change methodology was correlated with recidivism, it may be that the Christensen and Mendoza (1986) formula is too stringent to apply to this population. Reliable clinically significant change may in fact reflect actual

change in dynamic factors, which in turn are predictive of a reduction in recidivism, as long as reliable change methodology is not too restrictive to obscure genuine change.

It is important to note that significant correlations with recidivism for methods of reliable change were found to be positive. As described previously, correlations with recidivism were expected to be negative, demonstrating that greater pro social change would result in a decrease in recidivism. Positive correlations therefore suggest that greater pro social change is instead linked to an increase in recidivism. A possible explanation for this may be linked to Beggs and Grace (2011) findings regarding positive correlations between raw change scores and sexual recidivism. Here they concluded that this occurred due to raw change scores not taking in to account the pre-treatment scores. Because the psychometric battery consisted of measures that had minimum and maximum scores and many had items that were transparently worded (e.g., the ABCS), offenders who were more deviant at pre-treatment have more scope to demonstrate improvement on the scales when measured at post treatment. In this sense, whilst more deviant offenders are more likely to reoffend, they are also able to evidence greater treatment change. Considering this, reliable change measurements do not take in to account the pre-treatment scores and this therefore may account for the positive correlations found with sexual recidivism. Correlations with clinically significant change were in the expected direction as, in contrast to reliable change, they depend mostly on the post treatment scores. These findings suggest that reliable change methodology is not likely to be the most useful choice in regards to measuring treatment change in sex offenders.

Evaluating change across four risk domains and correlations with sexual recidivism

For both methods of calculating clinically significant change, results showed significant correlations with recidivism for all four dynamic risk factors (Allan et al., 2007). This again indicates that pro social change, as measured psychometrically is predictive of

81

reduced recidivism. However, again this is only for clinically significant change which may or may not reflect reliable or genuine change.

For residual change scores, results showed significant correlations for just two domains, Social Inadequacy (F1) and Anger/Hostility (F3). This suggests that BRCZ scores were not as strongly correlated as CSC and WCSC, however this may also be interpreted as residual change being a more stringent method of calculating or assessing change. These results are interesting in that they differ from Beggs and Grace (2011) findings regarding risk domains. Their results were significant for Anger/Hostility (F3) and Sexual Interests (F2) but not for Social Inadequacy (F1) and Pro-offending attitudes (F4). Essentially this study replicated the Beggs and Grace (2011) methodology in calculating residual change scores, however there were differences in the samples used. The current study used a larger sample (N=495 compared with N=218). The sample used in the current study was the same as that used by Allan et al. (2007) which included all offenders who entered Kia Marama from the programme's inception in 1989. The sample used by Beggs and Grace (2011) included participants who were released from the programme from 1993 to 2000. As expected, there were evaluations and enhancements of programme content and delivery over time, together with changes in the measures included in the psychometric battery administered. Alongside differences in numbers of participants, these factors contribute to dissimilarities overall between the samples and should be considered in the analysis and variances in the results.

In regards to reliable change methodology, again differences were apparent. No significant correlations were found between recidivism and RCSC across the four factors. This is to be expected given that these analyses were based on the same data set as the overall change score correlations. For Wakeling et al.'s (2013) less stringent method of calculating reliable change, only one domain, Social Inadequacy (F1), was found to be correlated with sexual recidivism. When Wakeling et al. (2013) excluded rapists from their sample and just

examined child molesters' data, they found significant associations between recidivism and two risk domains: Socio affective functioning and self-regulation. It is likely that the dimensions of 'socio affective functioning' and 'social inadequacy' show considerable overlap and may represent the same construct. Given that results were significant for these domains in two independent studies with child molesters, the implication is that pro social changes made in treatment regarding social inadequacy factors such as emotional control, self-esteem, depression and assertion, are specifically related to reductions in recidivism. It is likely that positive changes in these areas are directly related to reductions in specific dynamic risk factors relating to interpersonal functioning such as negative emotionality and relationships with others.

Percentages of change on all variables across all methods of calculating change

Overall, methods of calculating change provided differing assessments of the change showed by the sample, with measures of reliable change appearing to be considerably more stringent than measurements of clinically significant change. The most consistent result appears to be high rates of clinically significant change (for both methods of calculation) specifically for the measurement of distorted attitudes towards children and sex (ABCS). However, given the transparent nature of this scale and that reliable change scores were considerably lower, these results should be interpreted with caution. Whilst it appears encouraging that percentages for CSC are high, it does not necessarily mean that these reflect genuine treatment gains. As noted previously, a common limitation of studies incorporating self-report psychometrics is the issue of social desirability. Motivations to present oneself in a positive light are obvious for incarcerated sex offenders and therefore this should be taken in to consideration in the interpretation of results.

Regarding different formulas for calculating clinically significant change, it was apparent that the Wakeling method evidenced more change across the sample than this

study's method. Our CSC formula concludes that a person's score has moved into the functional population range at post-treatment, however it is unclear if the Wakeling calculation does this. To be only 1SD away from pre-treatment mean does not appear to be very difficult to achieve or necessarily guarantee pro-social functioning post treatment. This may account for the apparent levels of success regarding change calculated with the Wakeling methodology.

Wakeling et al. (2013) categorized their data in a number of ways. Initially they used the five categories of treatment change identified by Jacobson et al. (1999) (*deteriorated, unchanged, improved, recovered* and *already okay).* These categories incorporated all of the data collected and were based on clinically significant and reliable change calculations. Treatment change categories were then calculated for the SARN domains. For this, individuals had to be in the *recovered* or *already okay* category for at least half of the psychometric measures in that domain. This produced two treatment change categories; *change still required* and *change not required* which were then used to determine overall change. Individuals who fell in to the *change not required* category for at least two of the three domains were deemed to be *changed.* Wakeling et al. (2013*)* acknowledged that no statistical technique was employed to determine these categories. They also report that the Sexual Interests domain was omitted from the overall treatment change categorization (hence only measured as two out of three categories) due to only one third of the sample completing the Multi Phasic Sex Inventory (Nichols & Molinder, 1984). The categorization of scores and the omission of one of the four domains therefore involves the exclusion of some of the data collected for each individual subject. The current study used average change scores to calculate overall and individual change. In this way, all available data for each offender was included in all analyses with none being excluded by repeated categorisation. This may

account for the differences found between this and the present study in regards to correlations with recidivism.

It is important to note that the current study and Beggs and Grace (2011) included only sexual offences in the analysis of recidivism. Wakeling et al. (2013) found their base rate for sexual offences alone to be too low and therefore included sexual and violent recidivism. This could be a factor in explaining the differences in results. In similar analyses of treatment change, Barnett, Wakeling, Mandeville-Norden, and Rakestrow (2011) questioned the same concept and therefore re-examined their figures using just sexual recidivism data, concluding that there were no differences in the results. Similarly, there was a notable difference in follow up time between this study (13.1 years) and Wakeling et al. (2013) (minimum 2 years). However, overall, the studies appear to be comparable in regards to, for example, psychometrics administered, treatment modality and prison based treatment.

Limitations of the present study

This study's reliance on self-report psychometric scores is apparent and raises a number of challenges common to such analyses. Because norms are required for calculating cut off scores, the clinically significant change methodology based on cut off scores is therefore only as good as the norm on which calculations are based. Nunes et al. (2011) highlight the complexities of calculating representative norms, particularly for scales considered specific to sex offenders (for example deviant sexual interests) rather than assessing concepts applicable to the general population (for example depression or self-esteem scales). Similar issues are relevant in regards to establishing reliability coefficients required for the calculation of reliable change therefore Nunes et al. (2011) highlighted the importance of future work to aggregate norms and reliability coefficients. Keeling et al. (2006) also commented on the need for normative data to be based on appropriate samples

and further highlight the need for highly reliable instruments in studies based on psychometric assessment.

Clinically significant change methodology is clearly highly dependent on the quality of the measures used and therefore change calculations can only be considered to be as robust as the measures themselves. Transparent measures and self-report bias can significantly influence results based on psychometric evaluation. Beggs and Grace (2011) discuss the issue and highlight the obvious incentives incarcerated offenders may have to present themselves in a positive light. Their analysis of this issue found a strong linear relationship between pre-treatment and change scores with those who were most deviant at pre-treatment reporting the most pro social change post-treatment. Wakeling et al. (2013) similarly highlight this issue and advise the replication of their study incorporating a measure of response bias. A measurement of social desirability was not included in the current study and therefore a similar recommendation here would also be appropriate for future analysis of the current data.

Common to studies reliant on self-report psychometric information, missing data contributes to the limitations of the present study. Changes over time in the psychometric battery administered to programme participants, together with administration errors and refusal by offenders to complete pre or post testing, results in missing data for some subjects. Whilst ultimately inevitable in a study of this design, it has the potential to influence outcomes and therefore should be taken in to consideration in interpreting results.

Further limitations of the current study lie within the sample itself. The entry criteria for Kia Marama requires offenders to have child victims. The current data set therefore does not include rapists or other offenders with adult victims. The conclusions drawn here therefore are not applicable to a general population of sexual offenders. The Beggs and Grace (2011) sample (being a subset of the current sample) is also limited only to child molesters

however Wakeling et al. (2013) noted interesting findings in their results when they examined child molesters and rapists separately. They concluded that dynamic risk factors present for child molesters potentially had a stronger association with recidivism than those present for rapists. However they offer an alternative explanation in that child molesters were potentially more honest in their response to psychometrics. Replicating the current study with a mixed sample of offence type could therefore be a useful addition to research in this area.

Beggs and Grace (2011) analysed change as measured by three methods, including self-report psychometrics and two structured clinical rating scales. The current study did not employ an additional measure of change. The addition of a measure such as the Violence Risk Scale: Sex Offender version (VRS:SO; Olver et al., 2007a) would allow for calculations of concurrent validity and therefore provide the potential to validate of the change observed from psychometric scores.

Recommendations for future research

The purpose of the current study was not to determine if change was attributable to attending treatment, more just to identify whether change was observed. Therefore no control group was used in the present study. However this does not compromise our ability to compare different measures of change within the same sample.

Pre-treatment measurements of static risk levels (e.g. Static 99R) were not taken into consideration in the present analyses. Static risk was considered by both Beggs and Grace (2011) and Wakeling et al. (2013) and found to be relevant to significance of correlations with recidivism. As comparisons of change for varying risk levels were not a key aim of the current study, the extent to which risk level influences outcome is therefore unknown. Future work could incorporate static risk assessment and change analyses with the current sample.

Categorizing change scores is common to many studies seeking to analyse relationships between treatment change and recidivism (e.g. Barnett et al., 2011; Olver et al.,

2015; Wakeling et al., 2013). Employing categories such as Jacobson et al. (1999) five

change categories (deteriorated, unchanged, improved, recovered already okay) could add to

the overall understanding of the change achieved by the sample. Such categorisation allows

for more detailed analyses of the data, including identifying those who were already

functional pre-treatment and those who deteriorated. As noted in the results section,

numerous further categorisation has the potential to exclude valuable data, however future

work could incorporate change groupings to further categorise and explore the relationship

between clinically significant and reliable change and sexual recidivism. Precaution must be

taken in the generation of such categories. As noted by Olver, Beggs Christofferson, et al.

(2014), the language of clinically significant change categories may be inappropriate for this

specific population, for example labelling an individual as 'already okay' when they have

committed a serious sexual offence or deeming offenders to be 'recovered' when treatment is

frequently described as not to be considered a 'cure' for sexual offending. Similarly,

Wakeling et al. (2013) note caution in regards to using an arbitrary approach to the creation

of change categories. Statistical procedure, over and above clinical judgement, should be

considered essential in this regard in the generation of any methodology to categorise change

scores.

Conclusions

　　Overall, the present results support the use of self-report psychometrics in measuring

treatment change and predicting recidivism. Measures of clinically significant change were

found to be significantly correlated with recidivism. However, this was not necessarily true

when change was defined as both reliable and clinically significant. Our results suggest that

the Christensen and Mendoza (1986) reliable change index methodology is too stringent to

determine positive results regarding the utility of assessing treatment change as predictive of

recidivism. Residual change scores may an advantage as an assessment of change because

they assess an offender's change relative to other offenders. As the pre-treatment score is partialled out, the residual change scores may be useful for risk assessment, however as Beggs and Grace (2011) noted, their application may be too complex for clinicians. Clinically significant change and reliable change index scores have the potential to be more expedient for calculation by professionals within clinical settings, with the Wakeling et al. (2013) methodology perhaps providing the most useful method of calculating treatment change for sexual offenders based on psychometric self-reports.

References

Abel, Gene G, Gore, DavidK, Holland, C. L., Camps, Nancy, Becker, JudithV, & Rathner, Jerry. (1989). The measurment of the cognitive distortions of child molesters. *Annals of sex research, 2*(2), 135-152. doi: 10.1007/BF00851319

Abel, Gene G, Mittelman, Mary S, & Becker, Judith V. (1985). Sexual offenders: Results of assessment and recommendations for treatment. *Clinical criminology: The assessment and treatment of criminal behavior*, 191-205.

Alexander, Margaret A. (1999). Sexual offender treatment efficacy revisited. *Sexual Abuse: A Journal of Research and Treatment, 11*(2), 101-116. doi: 10.1007/BF02658841

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*(4), 347-367. doi: 10.1177/107906320701900402

Anderson, Ronald D., Gibeau, Dennis, & D'Amora, David A. (1995). The Sex Offender Treatment Rating Scale: Initial Reliability Data. *Sexual Abuse: A Journal of Research and Treatment, 7*(3), 221-227. doi: 10.1177/107906329500700305

Andrews, D, & Bonta, James. (2003). *The psychology of criminal conduct* (Vol. 3rd). Cincinnati, OH: Anderson Pub. Co.

Andrews, D, Bonta, JAMES, & Hoge, R. D. (1990). Classification for Effective Rehabilitation: Rediscovering Psychology. *Criminal Justice and Behavior, 17*(1), 19-52. doi: 10.1177/0093854890017001004

Andrews, D, Bonta, James, & Wormith, J Stephen. (2006). The recent past and near future of risk and/or need assessment. *Crime and delinquency, 52*(1), 7.

Bain, Barbara A., & Dollaghan, Christine A. (1991). The Notion of Clinically Significant Change. *Language, Speech, and Hearing Services in Schools, 22*(4), 264-270. doi: 10.1044/0161-1461.2204.264

Baker, Anthony W., & Duncan, Sylvia P. (1985). Child sexual abuse: A study of prevalence in Great Britain. *Child Abuse & Neglect, 9*(4), 457-467. doi: http://dx.doi.org/10.1016/0145-2134(85)90054-7

Bakker, Leon Wilhelm, & Westaway, Jane. (1998). *And There was Light-: Evaluating the Kia Marama Treatment Programme for New Zealand Sex Offenders Against Children*: Psychological Service, Department of Corrections.

Barbaree, H. E. (2005). Psychopathy, treatment behavior, and recidivism: an extended follow-up of Seto and Barbaree. *J Interpers Violence, 20*(9), 1115-1131. doi: 10.1177/0886260505278262

Barnett, Georgia D., Wakeling, Helen C., Mandeville-Norden, Rebecca, & Rakestrow, Janine. (2012). How Useful Are Psychometric Scores in Predicting Recidivism for Treated Sex Offenders? *International Journal of Offender Therapy and Comparative Criminology, 56*(3), 420-446. doi: 10.1177/0306624x11403125

Barnett, Georgia D., Wakeling, Helen, Mandeville-Norden, Rebecca, & Rakestrow, Janine. (2011). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime & Law, 19*(1), 85-110. doi: 10.1080/1068316X.2011.607820

Bauer, Stephanie, Lambert, Michael J., & Nielsen, Steven Lars. (2004). Clinical Significance Methods: A Comparison of Statistical Techniques. *Journal of Personality Assessment, 82*(1), 60-70. doi: 10.1207/s15327752jpa8201_11

Beck, A., & Shipley, B. E. (1989). U.S Bureau of Justice Statistics special report: Recidivism of prisoners released in 1983. Washington DC: U.S. Department of Justice.

Beck, A., Steer, R.A., & Brown, G, K. (1997). *Beck Depression Inventory - Second Edition - Manual*. San Antonia, TX: The Psychological Corporation

Beck, A., & Ward, C. H., Mendelson. M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of general psychiatry 4*, 561-571.

Becker, Judith V., & Murphy, William D. (1998). What we know and do not know about assessing and treating sex offenders. *Psychology, Public Policy, and Law, 4*(1-2), 116-137. doi: 10.1037/1076-8971.4.1-2.116

Beech, Anthony (1998). A Psychometric Typology of Child Abusers. *International Journal of Offender Therapy and Comparative Criminology, 42*(4), 319-339. doi: 10.1177/0306624x9804200405

Beech, Anthony , Fisher, Dawn D., & Thornton, David. (2003). Risk assessment of sex offenders. *Professional Psychology: Research and Practice, 34*(4), 339-352. doi: 10.1037/0735-7028.34.4.339

Beech, Anthony , Mandeville-Norden, Rebecca, & Goodwill, Alasdair. (2010). Comparing Recidivism Rates of Treatment Responders/Nonresponders in a Sample of 413 Child Molesters Who Had Completed Community-Based Sex Offender Treatment in the United Kingdom. *International Journal of Offender Therapy and Comparative Criminology*. doi: 10.1177/0306624x10387811

Beech, Anthony, Beckett, Richard C, & Fisher, Dawn. (1998). *STEP 3: An evaluation of the prison sex offender treatment programme*: Home Office London.

Beech, Anthony, Erikson, Matt, Friendship, Caroline, & Ditchfield, J. (2001). *A six-year follow-up of men going through probation-based sex offender treatment programmes*: Great Britain, Home Office, Research, Development and Statistics Directorate.

Beech, Anthony, & Ford, Hannah. (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 & Law, 12*(6), 685-701. doi: 10.1080/10683160600558493

Beech, Anthony, Friendship, Caroline, Erikson, Matt, & Hanson, R Karl. (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*(2), 155-167.

Beech, Anthony, & Hamilton-Giachritsis, CatherineE. (2005). Relationship Between Therapeutic Climate and Treatment Outcome in Group-Based Sexual Offender Treatment Programs. *Sexual Abuse: A Journal of Research and Treatment, 17*(2), 127-140. doi: 10.1007/s11194-005-4600-3

Beech, Anthony, Mandeville-Norden, R., & Goodwill, A. (2012). Comparing recidivism rates of treatment responders/nonresponders in a sample of 413 child molesters who had completed community-based sex offender treatment in the United kingdom. *Int J Offender Ther Comp Criminol, 56*(1), 29-49. doi: 10.1177/0306624X10387811

Beggs, Sarah. (2010). Within-treatment outcome among sexual offenders: A review. *Aggression and Violent Behavior, 15*(5), 369-379. doi: http://dx.doi.org/10.1016/j.avb.2010.06.005

Beggs, Sarah, & Grace, R. C. (2010). Assessment of dynamic risk factors: an independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sex Abuse, 22*(2), 234-251. doi: 10.1177/1079063210369014

Beggs, Sarah, & Grace, Randolph C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology, 79*(2), 182-192. doi: 10.1037/a0022900

Bonta, James. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.

Bowen, Erica, Gilchrist, Elizabeth, & Beech, Anthony R. (2008). Change in Treatment Has No Relationship With Subsequent Re-Offending in U.K. Domestic Violence Sample: A Preliminary Study. *International Journal of Offender Therapy and Comparative Criminology, 52*(5), 598-614. doi: 10.1177/0306624x08319419

Burt, Martha R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology, 38*(2), 217-230. doi: 10.1037/0022-3514.38.2.217

Caprara, Gian Vittorio. (1986). Indicators of aggression: The dissipation-rumination scale. *Personality and Individual Differences, 7*(6), 763-769. doi: http://dx.doi.org/10.1016/0191-8869(86)90074-7

Check, James Victor Patrick. (1984). *THE HOSTILITY TOWARD WOMEN SCALE.* (NK57719 Ph.D.), The University of Manitoba (Canada), Ann Arbor. Retrieved from http://search.proquest.com.ezproxy.canterbury.ac.nz/docview/303325436?accountid=14499

http://gr2tq4rz9x.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ProQuest+Dissertations+%26+Theses+A%26I&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&rft.genre=dissertations+%26+theses&rft.jtitle=&rft.atitle=&rft.au=Check%2C+James+Victor+Patrick&rft.aulast=Check&rft.aufirst=James+Victor&rft.date=1984-01-01&rft.volume=&rft.issue=&rft.spage=&rft.isbn=9780315098022&rft.btitle=&rft.title=THE+HOSTILITY+TOWARD+WOMEN+SCALE&rft.issn=&rft_id=info:doi/ ProQuest Dissertations & Theses A&I database.

Christensen, Larry, & Mendoza, Jorge L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*(3), 305-308. doi: http://dx.doi.org/10.1016/S0005-7894(86)80060-0

Davis, Mark H. (1980). A multidimensional approach to individual differences in empathy.

Descutner, Carol J., & Thelen, Mark H. (1991). Development and validation of a Fear-of-Intimacy Scale. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(2), 218-225. doi: 10.1037/1040-3590.3.2.218

Doi, Susan C., & Thelen, Mark H. (1993). The Fear-of-Intimacy Scale: Replication and extension. *Psychological Assessment, 5*(3), 377-383. doi: 10.1037/1040-3590.5.3.377

Draycott, Simon, Kirkpatrick, Tim, & Askari, Roxanna. (2011). An idiographic examination of patient progress in the treatment of dangerous and severe personality disorder: a reliable change index approach. *The Journal of Forensic Psychiatry & Psychology, 23*(1), 108-124. doi: 10.1080/14789949.2010.481720

Evans, Chris, Margison, Frank, & Barkham, Michael. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health, 1*(3), 70-72. doi: 10.1136/ebmh.1.3.70

Eysenck, Sybil B. G., & Eysenck, H. J. (1978). IMPULSIVENESS AND VENTURESOMENESS: THEIR POSITION IN A DIMENSIONAL SYSTEM OF PERSONALITY DESCRIPTION. *Psychological Reports, 43*(3f), 1247-1255. doi: 10.2466/pr0.1978.43.3f.1247

Fanslow, Janet L., Robinson, Elizabeth M., Crengle, Sue, & Perese, Lana. (2007). Prevalence of child sexual abuse reported by a cross-sectional sample of New Zealand women. *Child Abuse & Neglect, 31*(9), 935-945. doi: http://dx.doi.org/10.1016/j.chiabu.2007.02.009

Finkelhor, David. (1994). The international epidemiology of child sexual abuse. *Child Abuse & Neglect, 18*(5), 409-417. doi: http://dx.doi.org/10.1016/0145-2134(94)90026-4

Finkelhor, David, Shattuck, Anne, Turner, Heather A, & Hamby, Sherry L. (2014). The lifetime prevalence of child sexual abuse and sexual assault assessed in late adolescence. *Journal of Adolescent Health, 55*(3), 329-333.

Friendship, Caroline, Falshaw, Louise, & Beech, Anthony R. (2003). Measuring the real impact of accredited offending behaviour programmes. *Legal and Criminological Psychology, 8*(1), 115-127. doi: 10.1348/135532503762871282

Furby, Lita, Weinrott, Mark R., & Blackshaw, Lyn. (1989). Sex offender recidivism: A review. *Psychological Bulletin, 105*(1), 3-30. doi: 10.1037/0033-2909.105.1.3

Gallagher, Catherine A, Wilson, David B, Paul Hirschfield, MA, Coggeshall, Mark B, & MacKenzie, Doris L. (1999). Quantitative review of the effects of sex offender treatment on sexual reoffending. *Corrections Management Quarterly, 3*(4), 11.

Gambrill, Eileen D., & Richey, Cheryl A. (1975). An assertion inventory for use in assessment and research. *Behavior Therapy, 6*(4), 550-561. doi: http://dx.doi.org/10.1016/S0005-7894(75)80013-X

Hall, Gordon C. Nagayama. (1989). Self-reported hostility as a function of offense characteristics and response style in a sexual offender population. *Journal of Consulting and Clinical Psychology, 57*(2), 306-308. doi: 10.1037/0022-006X.57.2.306

Hall, Gordon C. Nagayama. (1995). Sexual Offender Recidivism Revisited: A Meta-Analysis of Recent Treatment Studies. *Journal of consulting and clinical psychology, 63*(5), 802-809. doi: 10.1037/0022-006X.63.5.802

Hansen, N. B., & Lambert, M. J. (1996). Clinical significance: An overview of methods. *Journal of Mental Health, 5*(1), 17-24.

Hanson, Karl, Bourgon, Guy, Helmus, Leslie, & Hodgson, Shannon. (2009). The Principles of Effective Correctional Treatment Also Apply To Sexual Offenders: A Meta-Analysis. *Criminal Justice and Behavior, 36*(9), 865-891. doi: 10.1177/0093854809338545

Hanson, Karl, & Bussiere, Monique T. (1998). Predicting relapse: a meta-analysis of sexual offender recidivism studies. *Journal of consulting and clinical psychology, 66*(2), 348.

Hanson, Karl, & Bussière, Monique T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*(2), 348-362. doi: 10.1037/0022-006X.66.2.348

Hanson, Karl, Gizzarelli, Rocco, & Scott, Heather. (1994). The Attitudes of Incest Offenders Sexual Entitlement and Acceptance of Sex with Children. *Criminal Justice and Behavior, 21*(2), 187-202.

Hanson, Karl, Gordon, Arthur, Harris, Andrew J. R., Marques, Janice K., Murphy, William, Quinsey, Vernon L., & Seto, Michael C. (2002). First Report of the Collaborative Outcome Data Project on the Effectiveness of Psychological Treatment for Sex Offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*(2), 169-194. doi: 10.1177/107906320201400207

Hanson, Karl, Harris, Andrew JR, Scott, Terri-Lynne, & Helmus, Leslie. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Vol. 5): Public Safety Canada Ottawa.

Hanson, Karl, & Morton-Bourgon, Kelly. (2004). Predictors of sexual recidivism: An updated meta-analysis 2004-02. *Ottawa, Canada: Public Safety and Emergency Preparedness Canada*.

Hanson, Karl, & Morton-Bourgon, Kelly E. (2005). The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies. *Journal of Consulting and Clinical Psychology, 73*(6), 1154-1163. doi: 10.1037/0022-006X.73.6.1154

Hanson, Karl, & Thornton, David. (2000). Improving Risk Assessments for Sex Offenders: A Comparison of Three Actuarial Scales. *Law and Human Behavior, 24*(1), 119-136. doi: 10.2307/1394432

Harris, Grant T., & Rice, Marnie E. (2003). Actuarial Assessment of Risk among Sex Offenders. *Annals of the New York Academy of Sciences, 989*(1), 198-210. doi: 10.1111/j.1749-6632.2003.tb07306.x

Hedderman, Carol, & Sugg, Darren. (1996). *Does treating sex offenders reduce reoffending?* : Home Office London.

Helmus, Leslie, Hanson, R. Karl, Thornton, David, Babchishin, Kelly M., & Harris, Andrew J. R. (2012). Absolute Recidivism Rates Predicted By Static-99R and Static-2002R Sex Offender Risk Assessment Tools Vary Across Samples: A Meta-Analysis. *Criminal Justice and Behavior, 39*(9), 1148-1171. doi: 10.1177/0093854812443648

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

Hudson, Stephen, Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic Risk Factors: The Kia Marama Evaluation. *Sexual Abuse: A Journal of Research and Treatment, 14*(2), 103-119. doi: 10.1177/107906320201400203

Hudson, Stephen, & Ward, Tony. (1996). Relapse prevention: Future directions. *Sexual Abuse: A Journal of Research and Treatment, 8*(3), 249-256. doi: 10.1007/BF02256640

Hugdahl, K, & Ost, LG. (1981). ON THE DIFFERENCE BETWEEN STATISTICAL AND CLINICAL-SIGNIFICANCE. *Behavioral Assessment, 3*(3-4), 289-295.

Jacobson, N, Follette, W, & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*(4), 336-352. doi: http://dx.doi.org/10.1016/S0005-7894(84)80002-7

Jacobson, N, Follette, William C., & Revenstorf, Dirk. (1986). Toward a standard definition of clinically significant change. *Behavior Therapy, 17*(3), 308-311. doi: http://dx.doi.org/10.1016/S0005-7894(86)80061-2

Jacobson, N, Follette, William C., Revenstorf, Dirk, Hahlweg, Kurt, Baucom, Donald H., & Margolin, Gayla. (1984). Variability in outcome and clinical significance of behavioral marital therapy: A reanalysis of outcome data. *Journal of Consulting and Clinical Psychology, 52*(4), 497-504. doi: 10.1037/0022-006X.52.4.497

Jacobson, N, Roberts, Lisa J., Berns, Sara B., & McGlinchey, Joseph B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67*(3), 300-307. doi: 10.1037/0022-006X.67.3.300

Jacobson, N, & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology, 59*(1), 12.

Jacobson, N, Wilson, Leanne, & Tupper, Cheryl. (1988). The clinical significance of treatment gains resulting from exposure-based interventions for agoraphobia: A reanalysis of outcome data. *Behavior Therapy, 19*(4), 539-554. doi: http://dx.doi.org/10.1016/S0005-7894(88)80022-4

Kazdin, A. E. (1977). Assessing the Clinical or Applied Importance of Behavior Change through Social Validation. *Behavior Modification, 1*(4), 427-452. doi: 10.1177/014544557714001

Kazdin, A. E. (2001). Almost Clinically Significant (p < .10): Current Measures May Only Approach Clinical Significance. *Clinical Psychology: Science and Practice, 8*(4), 455-462. doi: 10.1093/clipsy.8.4.455

Kazdin, A. E., & Wilson, G Terence. (1978). *Evaluation of behavior therapy: Issues, evidence, and research strategies*: Ballinger.

Keeling, Jenny A., Rose, John L., & Beech, Anthony R. (2006). An investigation into the effectiveness of a custody-based cognitive-behavioural treatment for special needs sexual offenders. *Journal of Forensic Psychiatry & Psychology, 17*(3), 372-392. doi: 10.1080/14789940600658293

Kirsch, Laura G., & Becker, Judith V. (2006). Sexual offending: Theory of problem, theory of change, and implications for treatment effectiveness. *Aggression and Violent Behavior, 11*(3), 208-224. doi: http://dx.doi.org/10.1016/j.avb.2005.08.001

Lambie, Ian, & Stewart, Malcolm W. (2012). Community Solutions for the Community's Problem: An Evaluation of Three New Zealand Community-Based Treatment Programs for Child Sexual Offenders. *International Journal of Offender Therapy and Comparative Criminology, 56*(7), 1022-1036. doi: 10.1177/0306624x11420099

Langan, Patrick A, Schmitt, Erica L, & Durose, Matthew R. (2003). Recidivism of sex offenders released from prison in 1994.

Lanterman, Jennifer L., Boyle, Douglas J., & Ragusa-Salerno, Laura M. (2014). Sex Offender Risk Assessment, Sources of Variation, and the Implications of Misuse. *Criminal Justice and Behavior, 41*(7), 822-843. doi: 10.1177/0093854813515237

Laws, D. R., & Marshall, W. L. (1991). Masturbatory reconditioning with sexual deviates: An evaluative review. *Advances in Behaviour Research and Therapy, 13*(1), 13-25. doi: http://dx.doi.org/10.1016/0146-6402(91)90012-Y

Lawson, J.S., Marshall, W.L., & McGrath, P. (1979). The Social Self-Esteem Inventory. *Educational and Psychological Measurement, 39*(4), 803-811. doi: 10.1177/001316447903900413

Levenson, Hanna. (1975). Multidimensional Locus of Control in Prison Inmates1. *Journal of Applied Social Psychology, 5*(4), 342-347.

Lösel, Friedrich, & Schmucker, Martin. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1*(1), 117-146. doi: 10.1007/s11292-004-6466-7

Malamuth, Neil M., Sockloskie, Robert J., Koss, Mary P., & Tanaka, J. S. (1991). Characteristics of aggressors against women: Testing a model using a national sample of college students. *Journal of Consulting and Clinical Psychology, 59*(5), 670-681. doi: 10.1037/0022-006X.59.5.670

Mandeville-Norden, Rebecca, Beech, Anthony, & Hayes, Elizabeth. (2008). Examining the effectiveness of a UK community-based sexual offender treatment programme for child abusers. *Psychology, Crime & Law, 14*(6), 493-512. doi: 10.1080/10683160801948907

Mann, Ruth, Webster, Stephen, Wakeling, Helen, & Marshall, William. (2007). The measurement and influence of child sexual abuse supportive beliefs. *Psychology, Crime & Law, 13*(5), 443-458.

Marlatt, G. Alan, & George, William H. (1984). Relapse prevention: Introduction and overview of the model. *British journal of addiction, 79*(3), 261-273.

Marques, J. K. (1988). The Sex Offender Treatment and Evaluation Project: California's new outcome study. *Annals of the New York Academy of Sciences, 528*(1 Human Sexual), 235-243. doi: 10.1111/j.1749-6632.1988.tb42077.x

Marques, J. K. (1999). How to Answer the Question: "Does Sex Offender Treatment Work?". *Journal of Interpersonal Violence, 14*(4), 437-451. doi: 10.1177/088626099014004006

Marques, J. K., Wiederanders, 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 Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment, 17*(1), 79-107. doi: 10.1177/107906320501700108

Marshall, W. L. (1971). A combined treatment method for certain sexual deviations. *Behaviour Research and Therapy, 9*(3), 293-294. doi: http://dx.doi.org/10.1016/0005-7967(71)90016-7

Marshall, W. L., Eccles, A., & Barbaree, H. E. (1993). A three-tiered approach to the rehabilitation of incarcerated sex offenders. *Behavioral Sciences & the Law, 11*(4), 441-455. doi: 10.1002/bsl.2370110410

Marshall, W. L., Marshall, L. E., & Ware, Jayson. (2009). Cognitive Distortions in Sexual Offenders: Should They All Be Treatment Targets? *Sexual Abuse in Australia and New Zealand, 2*(1), 21-33.

Matravers, Amanda. (2013). *Sex Offenders in the Community*: Routledge.

McGlinchey, Joseph B., Atkins, David C., & Jacobson, Neil S. (2002). Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy, 33*(4), 529-550. doi: http://dx.doi.org/10.1016/S0005-7894(02)80015-6

McLean, A., & Rush, C. . (1990). *Base rates and characteristics of convicted sexual offenders: A New Zealand study*. Unpublished manuscript. Department of Justice. Wellington, New Zealand.

Mendelson, Tamar, & Letourneau, Elizabeth J. (2015). Parent-Focused Prevention of Child Sexual Abuse. *Prevention Science*, 1-9.

Moore, Lucy. (2012). *A comparison of offence history and post release outcomes for sexual offenders against children in new zealand who did or did not attend the kia marama special treatment unit*. Unpublished thesis. University of Canterbury. Christchurch, New Zealand.

Moster, Aviva, Wnuk, Dorota W., & Jeglic, Elizabeth L. (2008). Cognitive Behavioral Therapy Interventions With Sex Offenders. *Journal of Correctional Health Care, 14*(2), 109-121. doi: 10.1177/1078345807313874

Neller, Daniel J. Frederick Richard I. (2013). Classification Accuracy of Actuarial Risk Assessment Instruments Classification Accuracy of Actuarial Risk Assessment Instruments. *Behavioral Sciences & the Law, 31*(1), 141-153. doi: 10.1002/bsl.2047

Nichols, H, & Molinder, I. (1984). Manual for the multiphasic sex inventory. *Tacoma, WA: Crime and Victim Psychology Specialists, 24*.

NOMS, Rehabilitation Services Group. *Women are deceitful scale*. Unpublished. Available from Rehabilitation Services Group, NOMS, 4th Floor, Clive House, 70 Petty France, London SW1H 9EX. UK.

Nowicki, S., & Duke, M. P. (1983). The Nowicki-Strickland Life-Span Locus of Control Scales: Construct Validation. In H. M. Lefcourt (Ed.), *Research with the locus of control construct: Vol. 2. Developments and Social Problems* (Vol. 2, pp. 9-49). New York: Academic.

Nunes, Kevin L., Babchishin, Kelly M., & Cortoni, Franca. (2011). Measuring Treatment Change in Sex Offenders: Clinical and Statistical Significance. *Criminal Justice and Behavior, 38*(2), 157-173. doi: 10.1177/0093854810391054

Nunes, Kevin L., Pettersen, Cathrine, Hermann, Chantal A., Looman, Jan, & Spape, Jessica. (2014). Does Change on the MOLEST and RAPE Scales Predict Sexual Recidivism? *Sexual Abuse: A Journal of Research and Treatment*. doi: 10.1177/1079063214540725

Olver, M, Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. P. (2014). Incorporating Change Information Into Sexual Offender Risk Assessments Using the Violence Risk Scale–Sexual Offender Version. *Sexual Abuse: Journal of Research and Treatment, 26*(5), 472-499. doi: 10.1177/1079063213502679

Olver, M, Beggs Christofferson, Sarah M., & Wong, Stephen C. P. (2015). Evaluation and Applications of the Clinically Significant Change Method with the Violence Risk Scale-Sexual Offender Version: Implications for Risk-Change Communication. *Behavioral Sciences & the Law, 33*(1), 92-110. doi: 10.1002/bsl.2159

Olver, M, Kingston, Drew A., Nicholaichuk, Terry P., & Wong, Stephen C. P. (2014). A psychometric examination of treatment change in a multisite sample of treated Canadian federal sexual offenders. *Law and Human Behavior, 38*(6), 544-559. doi: 10.1037/lhb0000086

Olver, M, Wong, S, Nicholaichuk, T, & Gordon, A. (2007a). The validity and reliability of the Violence Risk Scale-Sexual Offender version: assessing sex offender risk and evaluating therapeutic change. *Psychol Assess, 19*(3), 318-329. doi: 10.1037/1040-3590.19.3.318

Olver, M, Wong, Stephen C. P., Nicholaichuk, Terry, & Gordon, Audrey. (2007b). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*(3), 318-329. doi: 10.1037/1040-3590.19.3.318

Pithers, W. (1990). Relapse Prevention with Sexual Aggressors. In W. L. Marshall, D. R. Laws & H. E. Barbaree (Eds.), *Handbook of Sexual Assault* (pp. 343-361): Springer US.

Pithers, W, Marques, Janice K, Gibat, Cynthia C, & Marlatt, GA. (1983). Relapse prevention with sexual aggressives: A self-control model of treatment and maintenance of change. *The sexual aggressor: Current perspectives on treatment*, 214-239.

Plaud, Joseph J., & Bigwood, Scott J. (1997). A multivariate analysis of the sexual fantasy themes of college men. *Journal of Sex & Marital Therapy, 23*(3), 221-230. doi: 10.1080/00926239708403927

Proulx, Jean, Pellerin, Bruno, Paradis, Yves, McKibben, André, Aubut, Jocelyn, & Ouimet, Marc. (1997). Static and Dynamic Predictors of Recidivism in Sexual Aggressors. *Sexual Abuse: A Journal of Research and Treatment, 9*(1), 7-27. doi: 10.1177/107906329700900102

Przybylski, R. (2015). *Recidivism of Adult Sexual Offenders*. United States of America: Retrieved from https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=271086.

Quinsey, Vernon, Khanna, arunima, & malcolm, p. Bruce. (1998). A Retrospective Evaluation of the Regional Treatment Centre Sex Offender Treatment Program. *Journal of Interpersonal Violence, 13*(5), 621-644. doi: 10.1177/088626098013005005

Rice, Marnie E., & Harris, Grant T. (2003). The Size and Sign of Treatment Effects in Sex Offender Therapy. *Annals of the New York Academy of Sciences, 989*(1), 428-440. doi: 10.1111/j.1749-6632.2003.tb07323.x

Rice, Marnie E., & Harris, Grant T. (2011). Is androgen deprivation therapy effective in the treatment of sex offenders? *Psychology, Public Policy, and Law, 17*(2), 315-332. doi: 10.1037/a0022318

Russell, Dan, Peplau, Letitia A., & Cutrona, Carolyn E. (1980). The revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology, 39*(3), 472-480. doi: 10.1037/0022-3514.39.3.472

Scalora, Mario J., & Garbin, Calvin. (2003). A Multivariate Analysis of Sex Offender Recidivism. *International Journal of Offender Therapy and Comparative Criminology, 47*(3), 309-323. doi: 10.1177/0306624x03047003005

Schewe, Paul A., & O Donohue, William. (1996). Rape prevention with high-risk males: Short-term outcome of two interventions. *Archives of Sexual Behavior, 25*(5), 455-471.

Seager, James A., Jellicoe, Debra, & Dhaliwal, Gurmeet K. (2004). Refusers, Dropouts, and Completers: Measuring Sex Offender Treatment Efficacy. *International Journal of Offender Therapy and Comparative Criminology, 48*(5), 600-612. doi: 10.1177/0306624x04263885

Serin, RALPH C., Mailloux, DONNA L., & Malcolm, P. BRUCE. (2001). Psychopathy, Deviant Sexual Arousal, and Recidivism Among Sexual Offenders. *Journal of Interpersonal Violence, 16*(3), 234-246. doi: 10.1177/088626001016003004

Seto, michael c., & Barbaree, howard e. (1999). Psychopathy, Treatment Behavior, and Sex Offender Recidivism. *Journal of Interpersonal Violence, 14*(12), 1235-1248. doi: 10.1177/088626099014012001

Skelton, Alexander, Riley, David, Wales, David, & Vess, James. (2006). Assessing risk for sexual offenders in New Zealand: Development and validation of a computer-scored risk measure. *Journal of Sexual Aggression, 12*(3), 277-286. doi: 10.1080/13552600601100326

Slater, Christine, & Lambie, Ian. (2010). The highs and lows of working with sexual offenders: A New Zealand perspective. *Journal of Sexual Aggression, 17*(3), 320-334. doi: 10.1080/13552600.2010.519056

Speer, David C., & Greenbaum, Paul 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*(6), 1044-1048. doi: 10.1037/0022-006X.63.6.1044

Spielberger, C. D. (1983). *State Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists.

Spielberger, C. D. (1988). State-trait anger expression inventory research edition. *Professional manual. Odessa, FL: Psychological Assessment Resources*.

Stoltenborgh, Marije, van IJzendoorn, Marinus H., Euser, Eveline M., & Bakermans-Kranenburg, Marian J. (2011). A Global Perspective on Child Sexual Abuse: Meta-Analysis of Prevalence Around the World. *Child Maltreatment, 16*(2), 79-101. doi: 10.1177/1077559511403920

Tapp, James, Fellowes, Emma, Wallis, Nicola, Blud, Linda, & Moore, Estelle. (2009). An evaluation of the Enhanced Thinking Skills (ETS) programme with mentally disordered offenders in a high security hospital. *Legal and Criminological Psychology, 14*(2), 201-212. doi: 10.1348/135532508X336178

Thornton, David. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment, 14*(2), 139-153.

Thornton, David, Mann, Ruth, Webster, Steve, Blud, Linda, Travers, Rosie, Friendship, Caroline, & Erikson, Matt. (2003). Distinguishing and Combining Risks for Sexual and Violent Recidivism. *Annals of the New York Academy of Sciences, 989*(1), 225-235. doi: 10.1111/j.1749-6632.2003.tb07308.x

Tierney, David, & McCabe, Marita P. (2001). An evaluation of self-report measures of cognitive distortions and empathy among Australian sex offenders. *Archives of Sexual Behavior, 30*(5), 495-519.

Underhill, James, Wakeling, Helen C, Mann, Ruth E, & Webster, Stephen D. (2008). Male sexual offenders' emotional openness with men and women. *Criminal Justice and Behavior, 35*(9), 1156-1173.

Wakeling, H, & Barnett, Georgia. (2011). Measuring grievance thinking in sexual offenders: The revised Dissipation–Rumination Scale. *Journal of sexual aggression, 17*(3), 273-289.

Wakeling, H, Beech, Anthony R., & Freemantle, Nick. (2013). Investigating treatment change and its relationship to recidivism in a sample of 3773 sex offenders in the UK. *Psychology, Crime & Law, 19*(3), 233-252. doi: 10.1080/1068316X.2011.626413

Walters, Glenn D. (2006). Risk-Appraisal Versus Self-Report in the Prediction of Criminal Justice Outcomes: A Meta-Analysis. *Criminal Justice and Behavior, 33*(3), 279-304. doi: 10.1177/0093854805284409

Ward, Tony, & Beech, Anthony. (2006). An integrated theory of sexual offending. *Aggression and Violent Behavior, 11*(1), 44-63. doi: http://dx.doi.org/10.1016/j.avb.2005.05.002

Ward, Tony, & Gannon, Theresa A. (2006). Rehabilitation, etiology, and self-regulation: The comprehensive good lives model of treatment for sexual offenders. *Aggression and Violent Behavior, 11*(1), 77-94. doi: http://dx.doi.org/10.1016/j.avb.2005.06.001

Ward, Tony, Hudson, Stephen M., Marshall, William L., & Siegert, Richard. (1995). Attachment Style and Intimacy Deficits in Sexual Offenders: A Theoretical Framework. *Sexual Abuse: A Journal of Research and Treatment, 7*(4), 317-335. doi: 10.1177/107906329500700407

Ward, Tony, Melser, Joseph, & Yates, Pamela M. (2007). Reconstructing the Risk–Need–Responsivity model: A theoretical elaboration and evaluation. *Aggression and Violent Behavior, 12*(2), 208-228. doi: http://dx.doi.org/10.1016/j.avb.2006.07.001

Ward, Tony, & Stewart, Claire A. (2003). The treatment of sex offenders: Risk management and good lives. *Professional Psychology: Research and Practice, 34*(4), 353-360. doi: 10.1037/0735-7028.34.4.353

Watson, Teresa, & Vess, Jim. (2008). Short Term Reoffending by Child Victim Sex Offenders in New Zealand: A Comparison of Those with and without Extended Supervision. *Sexual Abuse in Australia and New Zealand, 1*(1), 44-52.

Webster, Stephen D., Mann, Ruth E, Thornton, David, & Wakeling, Helen C. (2007). Further validation of the short self-esteem scale with sexual offenders. *Legal and Criminological Psychology, 12*(2), 207-216.

Webster, Stephen D., Mann, Ruth E., Carter, Adam J., Long, Julia, Milner, Rebecca J., O'Brien, Matt D., . . . Ray, Nicola L. (2006). Inter-rater reliability of dynamic risk assessment with sexual offenders. *Psychology, Crime & Law, 12*(4), 439-452. doi: 10.1080/10683160500036889

Wilkinson, John. (2005). Evaluating Evidence for the Effectiveness of the Reasoning and Rehabilitation Programme. *The Howard journal of criminal justice, 44*(1), 70-85. doi: 10.1111/j.1468-2311.2005.00356.x

Wilson, G. (1978). The secrets of sexual fantasy. London: Dent.

Wise, Edward A. (2004). Methods for Analyzing Psychotherapy Outcomes: A Review of Clinical Significance, Reliable Change, and Recommendations for Future Directions. *Journal of Personality Assessment, 82*(1), 50-59. doi: 10.1207/s15327752jpa8201_10