# Dynamic Risk Factors in Violence Risk Assessment: A Multiple Time-Point Evaluation of the HCR-20 and START

**by**

**Adam Joseph Evans Blanchard**

B.A. (Hons.), Simon Fraser University, 2010

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Psychology
Faculty of Arts and Social Sciences



**SIMON FRASER UNIVERSITY**

**Fall 2013**

# Approval

| | |
|---|---|
| **Name:** | **Adam Joseph Evans Blanchard** |
| **Degree:** | **Master of Arts (Clinical Forensic Psychology)** |
| **Title of Thesis:** | ***Dynamic Risk Factors in Violence Risk Assessment: A Multiple Time-Point Evaluation of the HCR-20 and START*** |
| **Examining Committee:** | **Chair:** Dr. Rebecca Cobb<br>Associate Professor |

**Kevin S. Douglas**
Senior Supervisor
Associate Professor _____

**Kathleen L. Slaney**
Supervisor
Associate Professor _____

**Patrick Bartel**
External Examiner
Adjunct Faculty
Department of Psychology _____

**Date Defended/Approved:** November 18, 2013

## Partial Copyright Licence

## Ethics Statement


The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

# Abstract

The consideration of dynamic risk factors when conducting risk assessments is generally considered best practice. However, little empirical research can speak to intraindividual change over time in putatively dynamic risk factors included in violence risk assessments instruments, and even fewer studies can speak to whether this change is associated with violence. The present study investigated change on putatively dynamic scales included on the Historical-Clinical-Risk Management-20 (HCR-20) and the Short-Term Assessment of Risk and Treatability (START), using a prospective repeated measures design with a civil psychiatric and a correctional sample. Intraindividual change on these scales was seen in a notable proportion of the assessments. More change was seen on the HCR-20 when the reassessment interval was over two months compared to less than two months, whereas the proportion of change on the START scales was consistent across different reassessment intervals. As well, fluctuations on these scales were predictive of subsequent violence.



**Keywords**: dynamic risk factors; risk assessment; violence; HCR-20; START

## Dedication

First and foremost, I would like to dedicate this thesis to my wife for her continued love and support over the years. I would not have been able to complete this project without you. I would also like to dedicate this thesis to my parents for everything they have done to make me the person I am today. Finally, I would also like to dedicate this to all of my loved ones out there that hold a dear place in my heart. I could not have made it this far without everything you all have done to help and inspire me.

# Acknowledgements

The totality of this thesis is the result of not only my work, but also many others that have contributed in varying degrees to my education, training, and research. Without all of you, this project would not have been possible.

I would like to express the deepest thanks and appreciation to my amazing supervisor Dr. Kevin Douglas for his ongoing support and inspiration throughout my academic career. I am grateful for the opportunity to work with such an understanding, patient, skilled, and accomplished mentor, whose expertise has enriched my graduate career immensely. His contributions to this project from its inception to completion have greatly improved the quality of my work in countless ways.

I would also like to thank my secondary supervisor Dr. Kathleen Slaney who has contributed hugely to my ongoing pursuit of academic excellence. As a teacher and mentor, her knowledge and expertise have played a large role in guiding me through my undergraduate and graduate studies. Her insights and feedback on this project provided the utmost guidance in ensuring a clear and sound final product.

Finally, I would like to give a special thanks to all of the individuals involved in the current research projects. Too many people to name individually have invested countless hours toward the development and execution of this research. Thank you to all of the Douglas lab members for their efforts.

# Table of Contents

# List of Tables

# List of Figures

# List of Acronyms

| | |
|---|---|
| AUC | Area under the Curve (of Receiver Operating Characteristic) |
| GEE | Generalized Estimating Equations |
| HCR-20 | Historical-Clinical-Risk Management-20 |
| ICC | Intraclass Correlation Coefficient |
| LS/CMI | Level of Service / Case Management Inventory |
| LSI-R | Level of Service Inventory - Revised |
| MIC | Mean Inter-item Correlation |
| PCL-R | Psychopathy Checklist – Revised |
| PCL-SV | Psychopathy Checklist – Screening Version |
| RCI | Reliable Change Index |
| RNR | Risk-Need-Responsivity |
| SPJ | Structured Professional Judgment |
| START | Short-Term Assessment of Risk and Treatability |
| START:AV | Short-Term Assessment of Risk and Treatability: Adolescent Version |
| VRS | Violence Risk Scale |
| VRS-SO | Violence Risk Scale – Sexual Offender Version |

# Introduction

The field of violence risk assessment has seen significant advances in the past three decades (Douglas & Kropp, 2002; Lavoie, Guy, & Douglas, 2009; Monahan & Steadman, 2001; Otto & Douglas, 2010). Since its inception, it has been the subject of considerable scepticism and debate (Douglas & Kropp, 2002; Monahan & Steadman, 2001). Subsequently, considerable developments have occurred in understanding the multifaceted nature of violence, the process and conceptualization of risk assessment, the foundation of theoretically and empirically robust risk factors, and the development and validation of structured risk assessment instruments (Douglas & Kropp, 2002; Douglas & Skeem, 2005; Monahan et al., 2001; Quinsey, Harris, Rice, & Cormier, 2006). The original scepticism regarding practitioners' inability to predict violence has been counteracted by hundreds of independent studies and dozens of meta-analyses that support the reliability and validity of numerous risk assessment instruments (Campbell, French, & Gendreau, 2009; Guy, 2008; Mossman, 1994; Yang, Wong, & Coid, 2010).

A major development has been the unification of risk assessment with risk management, prevention, and treatment (Douglas & Skeem, 2005). The marriage of risk assessment and risk management, coupled with the pragmatic circumstance of violence-prone individuals increasingly being treated in community settings, has shifted the emphasis from one-time predictions to ongoing clinical responsibilities aimed at protecting the public (Dvoskin & Steadman 1994; Monahan, 1996; Skeem, Mulvey, & Lidz, 2000). Now, more than ever, practitioners are required to make decisions regarding a given individual's risk in various settings and contexts (Shah, 1978; Skeem et al., 2000). Related to the merger of risk assessment and management is the introduction of an understanding of risk as dynamic, changeable, or fluctuating (Douglas & Skeem, 2005; Lussier & Davies, 2011). Scholars originating from different theoretical frameworks have stressed the dynamic aspects of risk (e.g., Andrews & Bonta, 2010b; Douglas & Skeem, 2005; Hanson & Harris, 2000). These scholars have emphasized the need to more fully understand the changeable aspects of risk and risk factors.

In order to fully appreciate this dynamic conceptualization of risk, several important theoretical developments shall be reviewed, including the development of several risk factor typologies, the delineation of the ultimate purpose of risk assessments, and the development of several theoretical models of risk assessment and management. Subsequently, the focus will be shifted to examining empirical research relevant to the notion of dynamic risk. A review of this theoretical and empirical work highlights the fact that, although considerable advancements have been made, "the science of risk assessment currently lags far behind practice" with regards to dynamic risk (Douglas & Skeem, 2005, p. 367). That is, currently the vast body of theoretical knowledge regarding dynamic risk overshadows the dearth of empirical evidence. Accordingly, the present research was undertaken in order to investigate change and characteristics of change on theoretically dynamic scales included on two risk assessment instruments. Specifically, the present study sought to determine if in fact intraindividual change was seen on these putatively dynamic scales, describe some of the characteristics of any observed change, and investigate whether observed change in these scales is related to violence.

## Typologies of Risk Factors

A risk factor is defined as a measureable characterization of each individual that precedes the outcome of interest (violence in this case) and, when present, is statistically associated with an increase in the likelihood of the outcome over the base rate when the risk factor is not present (Hart, 2008; Kazdin, Kraemer, Kessler, Kupfer, & Offard, 1997; Kraemer et al., 1997). Several classification methods have been proposed for risk factors that are relevant to understanding dynamic risk, including separating risk factors based on their ability to vary over time, their degree or type of association with violence, and their relative association and interaction with other risk factors.

### *Changeability of Risk Factors*

Numerous scholars have highlighted the importance of classifying risk factors by their ability to fluctuate over time. According to this typology, there are two major types of risk factors: *static* (fixed) risk factors and *dynamic* (changeable) risk factors (Craissati

& Beech, 2003; Douglas & Skeem, 2005; Hanson, 1998; Heilbrun, 1997; Kazdin et al., 1997; Kraemer et al., 1997). Static risks factors are relatively time-invariant. They are not capable of change over time either spontaneously or when targeted with some intervention or management strategy. These risk factors can signal an aberrant developmental trajectory and a long-term propensity for violence (Hanson, 1998). That is, static risk factors are most useful for assessing long-term risk and separating individuals into groups based on risk (Craig, Browne, Stringer, & Beech, 2005). However, static risk factors are unable to indicate, or predict, when violence will occur and cannot be used to determine change in risk over time.

Dynamic risk factors are time-variant; they have the potential to change and, most importantly, are amendable through intervention (Craissati & Beech, 2003; Douglas & Skeem, 2005; Hanson, 1998). Scholars generally agree that identification of dynamic risk factors is essential as they present the best candidates for intervention (Andrews & Bonta, 2010a, 2010b; Douglas & Skeem, 2005; Hanson 1998). Dynamic risk factors have also been described in terms of their speed of change. That is, some dynamic factors may change more rapidly and frequently than others. Hanson and Harris (2000; see also Craissati & Beech, 2003; Douglas & Skeem, 2005; Hanson, 1998) describe *stable* and *acute* dynamic risk factors. Stable dynamic risk factors can change gradually over longer periods of time (i.e., months or years) but are unlikely to change quickly or over short periods. They are rather persistent characteristics and traits (Craig et al., 2005). On the other hand, acute dynamic risk factors are capable of changing abruptly and frequently (i.e., within weeks, days, or hours). However, few empirical investigations have tested the distinction between stable and acute dynamic risk factors.

The division of risk based on relative time-variance is often oversimplified. The line between a static risk factor and a stable dynamic risk factor is not always clear (Craig et al., 2005). For instance, psychopathy is often conceptualized as a pervasive and persistent set of features that do not change over the lifespan; however, it has also been shown to gradually respond to some forms of intervention (e.g., Edens, Skeem, & Kennealy, 2009; Salekin, 2002; Skeem, Monahan, & Mulvey, 2002). In many assessment instruments, psychopathy is considered a static risk factor, yet levels of psychopathic traits may fluctuate gradually over extended periods of time. Accordingly, some contemporary scholars in the area consider this typology along a continuum

ranging from completely fixed or static risk factors (e.g., age at first violent episode) to extremely variable (e.g., violent ideation) (Brown, Amand, & Zamble, 2009; Hanson & Harris, 2000; Quinsey, Jones, Book, & Barr, 2006).

## *Functional Relation of Risk Factors*

Scholars have also differentiated between risk factors based on their degree or type of association with violence. Kraemer and colleagues (1997; see also Dempster & Hart, 2002; Douglas & Kropp, 2002; Douglas & Skeem, 2005; Hart, 2008; Kazdin et al., 1997) developed a typology of risk factors based on the relative association of a characteristic with the outcome. A *correlate* is a characterization that is associated with the outcome, but no temporal or directional association is known. A *risk factor* is a correlate that has been shown to precede the outcome. That is, a risk factor precedes the outcome and is associated with an increased likelihood of the outcome. Risk factors are further divided based on their ability to change (fixed versus variable) and their functional relation to violence (risk marker versus causal risk factor). Accordingly, a *fixed risk marker* is a time-invariant variable that is not causally related to the outcome. A *variable risk marker* is a time-variant variable, but change on the variable does not result in corresponding change in the likelihood of the outcome. That is, any change on a variable risk marker is not causally linked with any change in the outcome. A characteristic that is capable of change and this change results in a corresponding change in the likelihood of the outcome is termed a *causal risk factor*. If no empirical evidence can speak to the stability of a risk factor over time, the general risk factor term is applied. If the risk factor has been shown to fluctuate over time, but no empirical evidence can speak to the effect of fluctuations in the risk factor on the likelihood of the outcome, the term *variable risk factor* is applied. Under this typology, causal risk factors are, thus, variables that (a) precede violence, (b) increase the likelihood of violence, (c) fluctuate over time either spontaneously or through intervention, and (d) are associated with corresponding fluctuations in the likelihood of violence. These risk factors are the most important to identify, as interventions targeted at these risk factors can result in a reduction in violence.

In addition, Hart (2008; see also Cooke & Michie, 2013; Douglas, Blanchard, & Henry, 2013) identified several ways in which a risk factor may be functionally related or

relevant to a given individual's perpetration of violence. This typology further classifies causal risk factors into three types. A *motivator* (referred to as a *driver* in Cooke & Michie's typology) is a risk factor that drives or directs an individual towards violence; it makes the individual perceive violence as an attractive or rewarding option. For instance, persistent relationship problems may lead an individual to believe that intimate partner violence is a good manner in which to express their feelings. A *disinhibitor* is a risk factor that loosens or dampens normal constraints on violent behaviour; the risk factor weakens the effects of normal societal restraints and prohibitions against violence. As an example, current intoxication or emotional instability may lessen the anxiety and reservation when dealing with an interpersonal conflict. A risk factor operating as an *impeder* interferes with the effectiveness of strategies designed to prevent violence. For example, negative attitudes towards authority may lead an individual to non-compliance with supervision and treatment designed to curb risk.

### *Interplay of Multiple Risk Factors*

In addition to the consideration of individual risk factors in isolation, it is important to consider the interrelations between risk factors and the role of these associations on the individual's risk level, as "the relations among risk factors and violence can be direct or indirect (mediated), unidirectional or bidirectional, or interactive (moderated)" (Douglas & Skeem, 2005, p. 368). Kraemer, Stice, Kazdin, Offord, and Kupfer (2001) discuss three factors that must be considered when dealing with a set (i.e., more than one) of risk factors. These factors are temporal precedence, correlation, and dominance. That is, which risk factor in the set precedes the other(s), whether the risk factors are correlated, and which risk factors alone or in combination will lead to the maximum predictive power. Analyzing these three factors assists in determining which risk factors are proxy risk factors for other variates, moderated or mediated by other factors, or independent or highly dependent on other factors.

## Purpose of Risk Assessments

In addition to considerable developments in the understanding and conceptualization of risk factors, the field has seen a significant shift in that the ultimate

goal of risk assessments has become the prevention, not prediction, of violence (Douglas & Kropp, 2002; Hart, 1998; Heilbrun, 1997). In a very influential article, Heilbrun (1997, see also Heilbrun, Nezu, Keeney, Chung, & Wasserman, 1998) distinguished between two models of risk assessment based on the underlying purpose. The *prediction* model is primarily concerned with making single time-point predictions of violence. In contrast, the *management* model is focused on ongoing assessments that direct intervention strategies to reduce violence. That is, the management model stresses the close relationship between risk assessment and risk management in that a comprehensive assessment should guide the intensity, selection, and targets of management strategies. Under the management model, a main focus is placed on the assessment of dynamic risk factors, especially those thought to be modifiable by targeted interventions.

The management model now dominates the field. Due to changes at multiple levels, including various legal, clinical, societal, and economic changes, the ultimate focal point of risk assessment has changed from the one-time prediction of violence to the ongoing assessment of risk in order to develop and implement strategies to reduce violence (Douglas, Cox, & Webster, 1999; Douglas & Skeem, 2005; Dvoskin & Heilbrun, 2001; Heilbrun, 1997; Hart, 1998; Otto, 2000; Skeem et al., 2000; Steadman, 2000; Steadman et al., 1993; Webster, Douglas, Belfrage, & Link, 2000). The process of risk assessments has shifted to "ongoing, day-to-day decisions" regarding the risk posed and management needed for a given individual (Steadman et al., 1993, p. 41). Accordingly, the main goal of risk assessment is now thought of as "the process of evaluating individuals to (1) characterize the risk they will commit acts of violence and (2) develop interventions to manage or reduce that risk" (Hart, 2001, p. 14; see also Hart, 2008).

Inherent in the management model of risk assessment, is a critical distinction between the notions of risk status and risk state (Douglas & Skeem, 2005; Heilbrun, Douglas, & Yasuhara, 2009; Skeem & Mulvey, 2002). *Risk status* refers to interindividual differences in risk level. Risk status identifies groups of individuals that pose a greater risk compared to others based on elevated levels of risk factors. As it is largely based on static risk factors, risk status is considered largely invariant over time. However, even amongst individuals identified as high risk based on risk status, a given individual's risk of violence "ebbs and flows" over time (Douglas & Skeem, 2005, p. 348).

*Risk state* indicates the intraindividual risk level of a given individual at a given moment in time. Risk state describes a particular individual's propensity for engaging in violence at a particular time based on fluctuations of the individual's biological, psychological, and social spheres (Douglas & Skeem, 2005). Although requiring the consideration of static risk factors, it is principally derived from consideration of the individual's current standing on dynamic risk factors. As such, a given individual's risk state is thought to fluctuate over time due to both external (e.g., treatment, supervision) and internal (e.g., learning, aging) forces, whereas risk status based on point estimates of static risk factors is relatively unchanging (Douglas & Skeem, 2005; Dvoskin & Heilbrun, 2001).

The introduction and consideration of risk state is relatively new to the field, and interest in this notion has been increasing since its introduction (Douglas & Skeem, 2005). Although there is still debate in the field, many scholars believe that a given individual's likelihood of engaging in violence waxes and wanes over time in relation to the individual's standing on various dynamic risk factors (Heilbrun et al., 2009). This change in focal point, to a management model and focus on risk state, has been accompanied by the development of risk assessment instruments to aid in the decision making process. The majority of violence risk assessment tools, especially those developed at an earlier time in the field, were constructed under the prediction framework (i.e., single time-point estimates of violence); however, more recent conceptualizations, theoretical models, and corresponding assessment tools incorporate dynamic risk (Andrews & Bonta, 2010a, 2010b; Douglas & Skeem, 2005; Grann et al., 2005; Webster, Douglas, Eaves, & Hart, 1997; Hart, 2008).

## Theoretical and Conceptual Models

Two prominent models currently in widespread use in the risk assessment field are particularly attuned to the current focus on dynamic risk. These models share many features and theoretical underpinnings, including focusing on dynamic risk factors, unifying risk assessment and management, and assessing risk state. They can be also seen as complementary in some ways as they were developed out of a primary focus in risk assessment (Structured Professional Judgment model) and risk management (Risk-Need-Responsivity model), respectively. However, as comprehensive models or

approaches, this simplification is not meant to disregard their unique characteristics and broad applicability.

## *Structured Professional Judgment Model*

The Structured Professional Judgment (SPJ) model of violence risk assessment is acutely in tune with the current understanding of the dynamic nature of risk. Developed in an effort to include the strengths of both actuarial (i.e., statistical or mechanical) and clinical (i.e., intuitive) approaches to prediction, the SPJ model sets out guidelines that reflect current empirical, theoretical, and clinical knowledge (Douglas & Kropp, 2002). The guidelines and procedures set out the necessary user qualifications, the relevant information to consider, the manner with which to gather and combine information, effective communication strategies, and methods for implementing intervention strategies (Douglas & Kropp, 2002; Douglas & Reeves, 2010).

Structure is imposed in the decision making process by (a) the inclusion of a fixed number (20 to 30) of operationally defined risk factors that must be considered in every case, (b) explicit coding rules for each risk factor, and (c) explicit instructions for the determination of final decisions about risk based on the number of risk factors that are present, the relevance of these risk factors to the individual being evaluated, any important interactions between these risk factors, how these risk factors manifest within the individual, and the nature and intensity of intervention strategies that are needed to mitigate the individual's risk (Douglas, Hart, Webster, Belfrage, & Eaves, 2008; Douglas, Ogloff, & Hart, 2003; Heilbrun et al., 2009).

At the same time, the SPJ model incorporates intraindividual differences, the idiographic nature of assessment, and links between risk assessment and management strategies (Heilbrun et al., 2009). By requiring professionals to determine the idiographic relevance and manifestations of each risk factor, the professional is forced to consider many of the risk factor typologies described above. Professional judgment is also required when arriving at a final risk judgment, which includes a determination of the appropriate risk management strategies to implement and the effect of these plans on the individual's risk state. As well, SPJ instruments include both static and dynamic risk factors in an attempt to connect and facilitate risk management strategies (Douglas &

Reeves, 2010). Guidelines are set out that specify the relevant time frame upon which to base ratings of dynamic factors and when to update the assessment of dynamic factors. In addition, the final risk judgment (low, moderate, or high) is considered a communicative aid to assist in the main goal of identifying relevant risk factors to target and implementing management strategies to mitigate risk (Douglas & Reeves, 2010; Heilbrun et al., 2009).

### *Risk-Need-Responsivity Model*

The Risk-Need-Responsivity (RNR) model of correctional treatment (i.e., risk management) describes three main principles for effective correctional programming (Andrews & Bonta, 2010a, 2010b; Andrews, Bonta, & Hoge, 1990; Andrews & Dowden, 2007; Dowden & Andrews, 2000). The risk principle sates that the intensity of services should be commensurate with the individual's risk level. That is, generally speaking, high-risk (status) individuals should receive high intensity services and low risk (status) individuals should receive minimal services. Individuals must undergo a comprehensive risk assessment in order to determine interindividual differences in risk level and assign them to levels of treatment intensity. The risk principle, thus, speaks to who should be targeted with services and the intensity of the services offered to different groups based on risk status. The need principle states that treatment must be directed at criminogenic needs. A criminogenic need is a dynamic risk factor that is functionally related to criminal behaviour (i.e., a causal risk factor of criminal behaviour). Criminogenic needs are the "targets of change" that management strategies focus upon (Andrews & Bonta, 2010a, p. 45). Thus, as with the risk principle, the need principle calls for a thorough assessment of the individual's risk (state) considering the different typologies of risk factors presented above. The need principle identifies which criminogenic needs must be targeted by some management strategies. The responsivity principle states that the treatment must be provided in a manner that is sensitive to the individual's learning ability and style. The responsivity principle involves using strategies that have been shown maximally effective generally and then individualizing the treatment according to characteristics of the particular individual. This principle speaks to how treatments should be implemented.

An important concept to clarify is a criminogenic need. Scholars often treat criminogenic need and dynamic risk factor as synonymous. For instance, Bonta (2002) states that "criminogenic needs are not simply 'any old needs' – they are dynamic risk factors" (p. 367). However, a criminogenic need is a risk factor that is capable of change and this change is associated with corresponding change in the likelihood of the outcome (Andrews et al., 1990). That is, the change on a criminogenic need is statistically related to change on the criterion, and the change shows incremental validity over and above baseline assessments (Andrews et al., 1990). Therefore, a criminogenic need refers only to the class of dynamic risk factors referred to by Kraemer and colleagues (1997) as causal risk factors. Dynamic risk factors constitute a broader category of variables that subsumes criminogenic needs.

## Theoretically Dynamic Risk Factors and Changeability

Significant theoretical advancements have been seen in the understanding and conceptualization of risk factors, the ultimate goals, purposes, and processes of risk assessments, and the prominent theoretical models of risk assessment and management (as described above). Although considerable advancements have been made in recent years, scholars have concluded that there is still insufficient guidance and direction available to practitioners who are tasked with ongoing monitoring, treatment, and decision-making (Belfrage & Douglas, 2002; Douglas et al., 2013; Douglas & Skeem, 2005). The understanding of dynamic risk and risk state "has, to date, been more conceptual and theoretical than empirical" (Belfrage & Douglas, 2002, p. 25). Consequently, the theoretical work needs to be thoroughly tested with empirical data. In particular, empirical investigations into the complex nature of change are needed. Intraindividual change in dynamic risk factors can vary in type, frequency, and rapidity. Change can be linear (relatively stable increase or decrease), quadratic (hasten in increase, and then decrease, or vice versa), cubic (hasten then slow in increase or decrease), or conform to higher order polynomial functions. Change can also fluctuate at different intervals (e.g., hours, days, weeks, months) on different patterns. Change may occur at different rates, in different directions, and at various wavelengths. As such, it is important to investigate empirically the changeability of theoretically dynamic risk factors, risk scales, and risk state.

## Empirical Research on Dynamic Risk

The present review focuses on the measurement of putatively dynamic risk factors[1] and risk state using a structured risk assessment model, specifically the SPJ or RNR model. A larger body of literature is available that pertains to investigations of the ability of single dynamic risk factors (operationalized as unitary constructs) to change over time, and a smaller body of literature reports on investigations of whether change on these constructs is associated with change in risk of violence. For instance, Douglas and Skeem (2005) provide a comprehensive review of several dynamic risk factors. They present research indicating that these constructs do in fact change over time and are in fact related to violence. However, the generalizability of these studies to risk assessment instruments is limited, as the ability to detect change will depend largely on the measurement and operationalization of the construct. These findings may hold true to the construct as operationalized in these studies, but this says little about risk factors as measured on standard risk assessment instruments. The manner in which individual constructs are operationalized and measured is often quite different from how they are operationalized and measured on various risk assessment instruments. Thus, the most appropriate manner in which to empirically investigate change in dynamic risk factors is to operationalize and measure them as they are used in clinical practice. That is, in order to investigate the changeability of dynamic risk factors over time and the relationship between observed change and change in the outcome, researchers should examine items and scales on structured risk assessment instruments. With that in mind, the available research may be separated by research design into single time-point evaluations (i.e., one assessment), dual time-point evaluations (i.e., two assessments), and multiple time-point evaluations (i.e., more than two assessments).

[1] Hence forth, items on risk assessment instruments that are theoretically considered to be dynamic and labeled as such on the instruments shall be referred to as *dynamic risk factors*. Scales on risk assessment instruments that consist solely of the above described items shall be referred to as *dynamic scales*, as they are labeled as such on the risk assessment instruments. However, as argued in the following sections, it is generally the case that the time-variance (or invariance) of these items or scales has not been adequately empirically investigated. As such, these items and scales should be considered putatively dynamic.

### *Single Time-Point Evaluations*

Currently there is a large body of research that has examined the relation between dynamic risk factors assessed at a single point in time and violence (Douglas & Skeem, 2005). Numerous reviews of the most predictive hypothetically dynamic risk factors are available (e.g., Craig et al., 2005; Craissati & Beech, 2003; Douglas & Skeem, 2005; Otto, 2000). Abundant independent empirical examinations have shown that dynamic risk factors included in structured risk assessment instruments are related to and predictive of violence. Both instruments that contain dynamic risk factors and dynamic scales (e.g., Daffern & Howells, 2007; Dolan & Fullam, 2007; Douglas et al., 2003; Douglas, Ogloff, Nicholls, & Grant, 1999; Douglas, Yeomans, & Boer, 2005; Simourd, 2004), as well as instruments consisting solely of dynamic risk factors (e.g., Chu, Thomas, Ogloff, & Daffern, 2011; Daffern et al., 2009; Grann et al., 2005; Nonstad et al., 2010) have been found to predict violence. Dynamic risk factors have been shown to be predictive of various indices and types of violence both in community (e.g., Douglas et al., 2003; Douglas et al., 1999; Douglas et al., 2005; Simourd, 2004; Wong & Gordon, 2006) and institutional settings (e.g., Belfrage, Fransson, & Strand, 2000; Chu et al., 2011; Daffern & Howells, 2007; Daffern et al., 2009; Dolan & Fullam, 2007). These findings have also been confirmed by several meta-analyses (e.g., Campbell et al., 2009; Guy, 2008; Yang et al., 2010).

In addition, studies have found that dynamic risk factors measured in single time-point evaluations add incremental validity to the statistical relationship between static risk factors and violence (e.g., Beech, Friendship, Erikson, M., & Hanson, 2002; Dempster & Hart, 2002). That is, dynamic risk factors add above and beyond the predictive validity of static risk factors. Studies have also shown that dynamic scales on risk assessment instruments add incrementally to the predictive validity of static scales on the same risk assessment instruments (e.g., Beggs & Grace, 2010; Vincent, Chapman, & Cook, 2011). Finally, dynamic scales and instruments have been shown to add unique predictive ability to the statistical association between well-validated risk assessment instruments and violence (e.g., Desmarais, Nicholls, Wilson, & Brink, 2012; Doyle & Dolan, 2006; Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012).

## *Dual Time-Point Evaluations*

As mentioned above, a larger body of research is available that has assessed individual dynamic risk factors operationalized in varying ways. Several studies have investigated whether dynamic risk factors (measured individually and in different manners) do in fact change over time and whether this change is related to crime or violence (e.g., Brown et al., 2009; Hanson & Harris, 2000; Jones, Brown, & Zamble, 2010; Odgers et al., 2009). Studies with adolescents and adults have also examined trajectories of offending longitudinally and included examinations of risk factors that are associated with different trajectories (e.g., Lussier & Davies, 2011; Nagin & Tremblay, 1999). As well, researchers have examined whether various dynamic constructs are able to predict treatment progress (e.g., Hildebrand & de Ruiter, 2012).

Several studies have examined change across two assessments using structured risk assessments tools. Many of these studies have focused on the Level of Service family of instruments. Andrews & Robinson (1984; as cited in Andrews, Bonta, & Wormith, 2010; see also Douglas & Skeem, 2005), with a sample of 57 non-disordered offenders on probation, assessed offenders using the Level of Service Inventory – Revised (LSI-R; Andrews & Bonta, 1995) at intake and six months later. They found that changes in overall risk level paralleled changes in recidivism rates at the six-month follow up. Motiuk, Bonta, and Andrews (1990; as cited in Andrews et al., 2010) found a similar pattern in a sample of 55 inmates assessed over a 12-month period using the LSI-R. Furthermore, Schlager and Pacheco (2011) investigated change on the LSI-R in a sample of 179 offenders supervised in the community in New Jersey. The LSI-R was administered two times approximately six months apart. Mean comparisons indicated an aggregate decrease in total LSI-R scores (27.1 to 24.0). Differences were also seen on eight of the 10 subcomponents, with aggregate decreases in mean scores on seven subcomponents and an aggregate increase in mean scores on one.

Holliday, Heilbrun, and Fretz (2012) examined 71 male correctional offenders undergoing treatment consistent with the RNR model in New Jersey. Participants stayed an average of 73.5 days in the program. They were assessed using the Level of Service / Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004),

upon entry to and exit from the program. The authors examined change on four criminogenic needs items on the LS/CMI: education/employment, family/marital, procriminal attitudes, and antisocial pattern. Based on mean comparisons, the authors found improvement in participant's overall LS/CMI score, with an average decrease of 1.3 points. They also found decreases on all four criminogenic need items. A larger degree of improvement was also seen for those classified in the highest level of need.

Raynor (2007) combined the results of two studies using dual time-point designs and the LSI-R. The first study included 948 offenders on probation in England and Wales, while the second study included 1380 offenders on probation in Jersey in the Channel Islands. The samples were divided into four groups based on their initial scores on the LSI-R (low or high starters) and direction of change (increase or decrease). Based on these subsamples ($n$ = 157 for the England and Wales sample; $n$ = 203 for the Jersey sample), Raynor reported higher recidivism rates for those whose scores increased regardless of their starting point compared to those whose scores decreased (67% versus 42%, respectively). The author concluded that this provided strong support for the fact that changes over time in the LSI-R scores are related to changes in the likelihood of recidivism. However, it should be noted that a substantial proportion of those whose scores decreased still went on to reoffend.

Additionally, Draycott, Kirkpatrick, and Askari (2012) analyzed data from 29 serious violent offenders in the dangerous and severe personality disorder program in England. Participants were assessed upon entry to and exit from the program, an average of 18 months apart. The assessments were made as part of routine practice using various measures including the Historical-Clinical-Risk Managment-20 (HCR-20, Webster et al., 1997). The HCR-20 is an SPJ instrument that includes two dynamic scales, the Clinical (C) scale and the Risk Management (R) scale, each with five items. The authors used reliable change indices to classify people based on their observed change as significant deterioration, deterioration, no change, improvement, or significant improvement. With regards to scores on the HCR-20 C scale, out of 29 participants, 1 participant was classified as deteriorated, 20 as no change, 7 as improvement, and 1 as significant improvement. With regards to the HCR-20 R scale, 1 participant was classified as deteriorated, 23 as no change, 4 as improvement, and 1 as significant improvement.

As well, one dual time-point study has examined change on the Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV; Nicholls, Viljoen, Cruise, Desmarais, & Webster, 2010). In an initial validation study of the START:AV, Viljoen and colleagues (2012) examined the performance of this instrument in a sample of 90 adolescents offenders on probation over a period of three months. At the group level, the authors found a mean decrease on the Vulnerability scale from baseline to follow-up, but no difference was seen in mean Strength scores between the two assessments. Intraindividual change was also examined using Reliable Change Indexes. The authors found that (reliable) change was seen in 17.5% of the Strength scores and 15.6% of the Vulnerability scores. As well, the authors reported that many of the final risk judgments (low, moderate, high ratings) changed over time (from 46.2% to 19.0% of ratings changed depending on the outcome).

Finally, several studies have investigated change on a specialized set of risk assessment instruments: the Violence Risk Scale (VRS; Wong & Gordon, 1999-2003) and Violence Risk Scale – Sexual Offender Version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003). These instruments assess change in a novel manner. Risk factors are first coded on a four-point scale (0 to 3) with higher scores indicating a stronger relationship with violence. Dynamic items that are coded as 2 or 3 are considered targets for treatment and subsequently rated on the stages of change model (Prochaska & DiClemente, 1986). At follow-up, these items are then recoded on the stage of change (i.e., precontemplation, contemplation, preparation, action, or maintenance). Progression from one stage of change to the next is then translated to a 0.5 reduction in the original four-point rating on each risk factor. Due to the unique nature of these assessments, these instruments are considered a class of their own compared to those reviewed above and below. Nevertheless, several dual time-point studies have found that ratings of this style do in fact change from pre-treatment to post-treatment and that change scores are predictive of subsequent violence (e.g., Beggs & Grace, 2011; Lewis, Olver, & Wong, 2013; Olver, Lewis, & Wong, 2013; Olver & Wong, 2011; Olver, Wong, Nicholaichuk, & Gordon, 2007).

## *Multiple Time-Point Evaluations*

Multiple time-point evaluations typically provide the most sound and robust opportunity to examine the dynamic nature of risk. Five studies have examined change on structured assessment instruments using multiple time-points. Belfrage and Douglas (2002) examined 150 male forensic psychiatric patients at two maximum-security forensic clinics in Sweden. A subgroup of this sample (*n* = 70) was assessed three times approximately six months apart using the HCR-20. In the subsample, examined longitudinally using mean comparisons tests, they found that mean scores decreased on the C and R scales (rated for continued inpatient care, but not outpatient care). They also saw mean score decreases on five individual items: negative attitudes (C2), unresponsive to treatment (C5), plans lack feasibility (R1), exposure to destabilizers (R2), and noncompliance with remediation attempts (R4).

Neves, Goncalves, and Palma-Oliveira (2010), using a sample of 158 correctional offenders on parole or probation in Portugal, examined change on the HCR-20 across three assessments over 13 months. They found no change in mean total scores over the three assessments. The only scale differences seen were increases on the R scale between the first and second assessment, and the first and third assessment. The only individual item to evidence change was on the R scale, stress (R5), which increased after the first assessment and then decreased after the second assessment. These authors also examined the predictive validity of the HCR-20 for general recidivism across the three assessments. They found that total and scale scores were predictive of recidivism. Conversely, they did find that recidivists' scores tended to increase across the assessment, while non-recidivists' scores did not increase.

Douglas, Strand, and Belfrage (2011) examined change on the HCR-20 C scale across four assessments each six months apart in a sample of 174 forensic psychiatric inpatients. The perpetration of violence was also determined at each interval. They found a linear decrease in total C scale scores over the four assessments. Using cluster analysis, they found five groups with different patterns of change: three groups decreased from different starting points, one group remained low across assessments, and one group rose then tapered off. As well, using repeated-measures ANOVA, they

found that observed change in C scale scores from the first assessment to the second assessment was predictive of subsequent violence.

As well, using a sample of 30 male forensic psychiatric inpatients and a pseudo-prospective research design, Wilson, Desmarais, Nicholls, Hart and Brink (2013) coded both the HCR-20 and the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009) from file information every three months for a one-year period. Dynamic risk factors were found to be predictive of institutional violence. Most importantly, using survival analyses, the authors found that change in dynamic risk scales on the HCR-20 was associated with future violence; moreover, this association was present even after controlling for the static items on the HCR-20. With regard to the START, the authors found that the changes in the vulnerability total score were associated with future violence even after controlling for the static items on the HCR-20. However, changes in the strength total score were not related to future violence.

Additionally, Michel and colleagues (2013) examined the performance of the HCR-20 using a prospective repeated measures design. Participants were assessed five times each, six months apart. Using a sample of 248 males with schizophrenia that included both forensic and general psychiatric patients, Michel et al. found that many of the C and R items changed at a group level over time. That is, they found group level differences across the five assessments. Mean scores on both the C and R scales changed over the five assessments. However, at the item level, the specific items that changed differed between the two subsamples. For the forensic patients, change was seen in two of the C items and four of the R items. For the general psychiatric patients, change was seen on only one C item and four R items. Michel and colleagues also reported the percentage of individual's scores that changed between any of the five assessments over the two-year period, with individual items changing in 42% (Negative Attitudes) to 83.0% (Unresponsive to treatment) of participants. Finally, the authors found that change in three of the C items (negative attitudes, impulsivity, and unresponsive to treatment) and change in three of the R items (plans lack feasibility, lack of personal support, and noncompliance with remediation attempts) was associated with future aggressive behaviours.

## Current Understanding of Dynamic Risk

Upon reviewing empirical investigations of dynamic risk, it is apparent that the majority of available empirical research can only speak to the ability of dynamic risk factors to explain interindividual differences in the perpetration of violence. There is a vast body of research that can speak to the general ability of (hypothetically) dynamic risk factors to predict violence, and to add incrementally to the predictive ability of static risk factors. It is still the case that the majority of empirical studies "use single time-point estimates of a putatively dynamic construct" (Douglas & Skeem, 2005, p. 356; see also Lussier & Davies, 2011; Olver & Wong, 2011). These studies form a solid foundation for the inclusion of dynamic risk factors in assessment instruments, yet they cannot be considered to address the central issue of dynamic risk.

There is scant empirical work to date that has investigated the changeability of dynamic risk factors on structured risk assessment instruments, even less empirical work has investigated intraindividual change on dynamic risk factors, and less research still has investigated whether this change is associated with or predictive of change in the occurrence of violence (Douglas & Skeem, 2005; Skeem, Mulvey, Lidz, Gardner, & Schubert, 2002). Several studies have determined that dynamic risk factors and scales on structured risk instruments have the potential to change over time (e.g., Belfrage & Douglas, 2002; Douglas et al., 2011; Draycott et al., 2012; Holliday et al., 2012, Michel et al., 2013; Neves et al., 2010; Viljoen et al., 2013). However, the results from these studies are far from conclusive. Change was seen on some items but not others and contradictory results were obtained. Notably, many of these studies found that some putatively dynamic risk factors did not change over extended periods of time, calling into question whether these risk factors are in fact dynamic, as they are currently operationalized (Belfrage & Douglas, 2002; Draycott et al., 2012; Neves et al., 2010). These studies also rarely examined the characteristics of any change that was found (i.e., frequency, speed, nature). Change was seen in as little as 74 days (Holliday et al., 2012) to as long as 18 months (Draycott et al., 2012). Studies that examined the same measure, the HCR-20, reported results ranging from change in one dynamic item (Neves et al., 2010) to five dynamic items (Belfrage & Douglas, 2002), to no or minute

change at the scale level (Draycott et al., 2012; Neves et al., 2010) to large and diverse change at the scale level (Douglas et al., 2011).

In addition, some studies used inappropriate statistical analyses to investigate intraindividual change. Group based mean comparisons were used in four studies (Belfrage & Douglas, 2002; Holliday et al., 2012, Neves et al., 2010; Schlager & Pacheco, 2011), which may mask change occurring at the individual level. Change at the aggregate or group level cannot speak directly to the central issue of whether dynamic risk factors and dynamic scales are capable of measuring intraindividual change over time. More appropriate analyses must consider change at the idiographic level, as some people may deteriorate or improve at different frequencies and speeds. For instance, Douglas and colleagues (2011) found five general patterns of change in their sample of forensic psychiatric inpatients. Moreover, Viljoen and colleagues (2012) found no group level, mean change from one assessment to the next, yet a notable proportion of the sample (15.6% to 17.5%) did display reliable intraindividual change on the same scales. Researchers must make clear the ultimate question they wish to answer and select the analysis that allows for the clearest answer.

From the available research it is impossible to draw any firm conclusions about the ultimate issue of whether change in dynamic risk factors is associated with violence. Douglas and colleagues (2011), Wilson and colleagues (2013), Michel and colleagues (2013) and several dual-time point studies with the LS instruments (e.g., Raynor, 2007) found that changes in dynamic risk factors were predictive of violence providing preliminary empirical evidence to confirm the extant theoretical work described above. As well, several studies have found that change as assessed in a unique manner on the VRS instruments is predictive of violence (e.g., Beggs & Grace, 2011; Lewis et al., 2013; Olver et al., 2013; Olver et al., 2007). One the other hand, Neves and colleagues (2011) did not find any differences in predictiveness across assessment, yet they did find weak evidence for differential change in recidivists compared to nonrecidivists. Thus, in general it is possible to tentatively conclude that some of the dynamic risk factors on these structured assessment tools are in fact dynamic, or are variable risk factors according to Kraemer and colleagues (1997) typology. However, some of these items may more accurately be described as general risk factors, as insufficient evidence has addressed their ability to fluctuate over time. Finally, it may be that some of these items

are in fact causal risk factors, or criminogenic needs, as change in the risk factors has been shown to be associated with change in the outcome, yet this conclusion requires far more definitive evidence.

## Current Research: Multiple Time Point Evaluation of the HCR-20 & START

Based on the aforementioned review and conclusions, the current research was undertaken to investigate the changeability of clinical ratings of dynamic scales on the HCR-20 and START over time using a multiple time-point longitudinal research design. The present study aims to advance the field by addressing empirically some of the fundamental questions regarding change and characteristics of change in putatively dynamic scales over time. Specifically, the current project will attempt to answer the following research questions regarding change on dynamic scales:

1. Do ratings of putatively dynamic risk scales change over multiple assessments at the individual level? Ultimately the primary question is that of whether or not change is seen in clinical ratings of putatively dynamic risk scales from one assessment to the next in a given individual. The risk factors included on the HCR-20 and START include dynamic constructs in an attempt to capture intraindividual change over multiple assessments, yet the empirical support for the intraindividual changeability of these risk factors is lacking and inconclusive.
2. Do ratings of putatively dynamic risk scales change differently depending on the length of the reassessment period? The second question of interest addresses the rate of change in these ratings. That is, if in fact change is seen across assessments, this intraindividual change may occur at different rates. Thus, more or less change may be seen when the length of reassessment interval is altered. Clinically this question has implications for the appropriate reassessment interval in various correctional and forensic settings (i.e., monthly, bimonthly, etc.).
3. What is the type or shape of change across multiple assessments at the group level? Following from the previous questions, it is important to determine the type of change that is seen across multiple assessments. The inclusion of six assessment periods allows for the investigation of higher order change functions. For instance, it is possible to determine whether change is linear, quadratic, or cubic in nature, and what direction any observed change takes (i.e., increase, decrease, both).

4. Do people change differently with respect to putatively dynamic risk scales? That is, is the type of change consistent across the entire sample, or are there identifiable groups that change differently on the assessment instruments? As some previous research has indicated (Douglas et al., 2011), there may be groups of individuals that change at different rates and according to different patterns.
5. Is change in ratings of putatively dynamic risk scales associated with future violence? Finally, if in fact change is seen on putatively dynamic scales on the HCR-20 and START, the next question of interest addresses whether or not this change is associated with future violence. From a clinical standpoint, if change is associated with violence, then interventions targeted at specific risk factors may be able to reduce the likelihood of future violence, which is the ultimate point of risk assessment (i.e., preventing future violence).

# Method

## Study Participants

A total of 235 participants were recruited for the present study from two subsamples of interest: a civil psychiatric sample and a correctional sample. The sample characteristics of each subsample are discussed in turn.

### *Civil psychiatric sample*

This sample consisted of 149 psychiatric inpatients who were admitted either voluntarily or involuntarily to the acute stay ward of a large hospital in British Columbia (Royal Columbian Hospital). Participants were eligible for inclusion in the study if they were (a) between the ages of 19 and 50[2], (b) fluent in English, (c) planning to reside in the greater Vancouver area, and (d) not diagnosed with mental retardation. The psychiatrists on the ward provided researchers with the names of patients who are eligible for participation. These patients were then approached by the research personnel, told about the study, asked if they would like to participate, and guided through the consent procedures. Any patient who a researcher deemed incompetent to provide informed consent, as assessed by multiple-choice questions in the consent form, was not included in the study.

The sample was split evenly across gender (51.7% male). With a mean age of 33.80 years (*SD* = 10.18), the majority of the sample was Caucasian (79.2%) or Asian (10.7%). Most of the participants had never been married (62.4%), while smaller portions were currently married (18.1%) or divorced (10.6%). As well, 24.3% of the participants were currently involved in a romantic relationship and 32.4% had at least

[2] This inclusion criterion was relaxed as the study progressed resulting in an age range from 19 to 61 in the present sample, with a total of 5 participants over the age of 50.

one child. The mean number of years of education was 12.79 (*SD* = 2.30), yet a minority of the sample had never completed high school (27.5%). With regard to their living situation prior to being admitted to the hospital, 36.2% were living with family, 30.9% were living alone, 14.8% were living in subsidized housing, and 4.7% were homeless. As well, approximately one third of the sample (32.1%) had been homeless (no fixed address) at some point prior to the baseline assessment. The patients were predominantly admitted to the psychiatric ward involuntarily (77.2%). With regard to their lifetime mental health diagnoses, psychotic disorders were the most common (44.9%), followed by bipolar disorder (39.5%), a depressive disorder (38.1%), some type of substance disorder (33.3%), and finally an anxiety disorder (21.3%). A majority of the participants had prior psychiatric hospitalizations (68.7%), with a mean of 2.65 (*SD* = 3.04) hospitalizations. A notable portion of the sample had previously been incarcerated (17.8%), with 36.2% coming into formal contact (arrest, charge, or conviction) with the criminal justice system for a non-violent offence and 15.9% having a formal contact for a violent offence.

### *Correctional sample*

This sample consisted of 86 correctional offenders who were serving sentences at one of four correctional institutions or four probation offices in the Lower Mainland of British Columbia (Fraser Regional Correctional Centre, North Fraser Pretrial Centre, Surrey Pretrial Services Centre, Allouette Correction Centre for Women, Burnaby Probation & Family Court, Abbotsford Community Corrections, Tri-Cities Community Corrections, and Vancouver Intensive Supervision Unit). In order to be serving a sentence in the custody of a provincial corrections institution or probation office in British Columbia, offenders must have been sentenced to a jail term of two years minus a day, or less. Eligibility criteria for this sample are identical to those for the psychiatric sample with the following two additional criteria[3]. Offenders recruited from a correctional institution must have been serving a minimum jail sentence of one month and must have been within one month of release. Offenders on probation must have been under

[3] The inclusion criterion regarding age was relaxed as the study progressed resulting in an age range from 19 to 61 in the present sample, with a total of 2 participants over the age of 50.

supervision for at least one month and be under supervision for at least 6 more months. Roughly equal numbers of mentally disordered and non-mentally disordered offenders were recruited, with mentally disordered offenders being classified as those with a current or lifetime diagnosis of any mood or psychotic disorder. Recruitment procedures for this sample are identical to those in the psychiatric sample, with the exception that corrections personnel referred the names to researchers and some sites directly advertised to the offenders who then referred themselves.

Both males (51.2%) and females were included in the sample. The majority of the sample was Caucasian (73.8%) or Aboriginal (15.5%), with a mean age of 34.65 years (*SD* = 8.41). Most of the sample had never been married (64.7%) while some were currently married (12.9%) or divorced (10.6%). As well, 37.3% of participants in this subsample were currently involved in a romantic relationship and 45.3% had at least one child. The mean number of years of education was 11.22 (*SD* = 2.06), with almost half of the sample having never completed secondary school (48.2%). With regard to their living situation prior to being sentenced for this index offence, 25.6% were homeless, 23.2% were living alone, 22.0% were living in subsidized housing, and 12.2% were living with family. A large portion of the sample (75.9%) had been homeless (no fixed address) at some point prior to the baseline assessment. As well, the sample consisted of both mentally disordered offenders (60.5%) and non-mentally disordered offenders. With regard to their lifetime mental health diagnoses, some type of substance use disorder was the most common (59.4%), followed by a depressive disorder (34.4%), an anxiety disorder (31.3%), a bipolar disorder (25.0%), and finally a psychotic disorder (21.9%). Almost half of the participants had previously been hospitalized for mental health reasons (44.0%), with a mean of 1.49 (*SD* = 2.86) psychiatric hospitalizations.

The sample was roughly equally split between offenders serving provincial jail sentences (54.7%) and those serving community sentences (45.3%). Nearly half of the participants were serving sentences that include both a jail term followed by a community probation term (49.4%), while others were serving only community probation term (30.1%). Many of the participants had served a previous term of incarceration (67.5%). In terms of the participants' index offences, 57.0% were for crimes against property, including 32.6% serving sentences for theft, 40.7% were for crimes against

persons, including 19.8% for some for of assault, and 11.6% for drug related offences, including 7.0% for trafficking offences.

## Measures

### *Historical-Clinical-Risk Management-20 (HCR-20)*

The HCR-20 (Webster et al., 1997) is one of the first violence risk assessment protocols developed under the SPJ model. It is intended to be used in order to facilitate assessments of risk for violence in civil psychiatric patients, forensic psychiatric patients, and criminal offenders both mentally disordered and not (Douglas, 2008). Specific settings and contexts in which the HCR-20 is used frequently include release and admission decision-making, as well as monitoring the risk of incarcerated, institutionalized, or community supervised individuals (Douglas & Reeves, 2010).

The HCR-20 consists of 20 risk factors grouped into three domains: Historical, Clinical and Risk Management. As seen in Appendix A, the measure contains 10 Historical, mostly static risk factors, 5 Clinical, putatively dynamic risk factors related to current functioning, and 5 Risk Management, potentially dynamic risk factors related to future considerations. Of note, in order to rate an item on the HCR-20, the evaluator is also required to complete either the *Psychopathy Checklist – Revised* (PCL-R; Hare, 1991; used with the correctional sample) or the *Psychopathy Checklist – Screening Version* (PCL-SV; Hart, Cox, & Hare, 1995; used with the civil psychiatric sample). In order to code the HCR-20, information from five general categories is used: file information, an interview with the individual being assessed, psychological and other tests, direct observation, and interviews with collateral individuals (Douglas, Hart et al., 2008; Douglas & Reeves, 2010).

Each risk factor is coded from 0 to 2, a coding practice that is common in SPJ instruments. A rating of 0 indicates that the item does not apply or is absent. A rating of 1 indicates that the item is possibly present or present only to a limited degree. A rating of 2 indicates that the item is definitely present. After scoring each item, final risk ratings of low, moderate or high risk for violence are made. In making these final risk judgments, assessors are encouraged to consider the number of risk factors present, the

relevance of each of these risk factors to the case at hand, and the types and intensity of services required to mitigate this person's need. For the current study, only the dynamic scales[4] (Clinical and Risk Management) were investigated and total scores on each of these scales were calculated for the subsequent analyses. In practice, the summation of numerical scores on this instrument is not recommended, yet this is acceptable for research purposes.

More than 100 examinations of this protocol's validity have been conducted (e.g., Douglas, Blanchard, Guy, Reeves, & Weir, 2010; Guy, 2008). Previous research has found the HCR-20 to have good predictive validity and inter-rater reliability. Both narrative (Douglas, Guy et al., 2008) and meta-analytic (Guy, 2008) reviews of the instrument have confirmed these findings. In terms of predictive validity, a recent meta-analysis found that the HCR-20 summary risk ratings were predictive of violence with an average AUC of .76, compared to an AUC of .73 for the total score (Guy, 2008). Additional meta-analyses have found that the HCR-20 on average across studies produces some of the largest effect sizes compared to other structured risk assessment instruments (e.g., Campbell et al., 2009; Yang et al., 2010). The majority of studies regarding inter-rater reliability have found the intraclass correlation coefficients (ICC) to be greater than .80 (Douglas & Reeves, 2010). Across 36 studies, Douglas and Reeves (2010) report that the inter-rater reliability to be in the good to excellent range (ICC = 0.67 to 0.95).

### *Short-Term Assessment of Risk and Treatability (START)*

The START (Webster et al., 2009) is a risk assessment instrument developed under the SPJ model that is meant to facilitate assessments of risk across seven domains: violence to others, suicide, self-harm, victimization, substance use, unauthorized absences, and self-neglect. However, the current project will only focus on the risk of violence to others. The measure includes 20 putatively dynamic risk factors (seen in Appendix B) that are rated as both strengths and vulnerabilities. The rating of strengths and vulnerabilities are independent in that a person may have any combination

[4] The dynamic scales on the HCR-20 contain putatively dynamic risk factors. Both the risk factors and scales are, thus, considered to be theoretically dynamic in nature.

of strength and vulnerability ratings on a single given risk factor. As with the HCR-20, each risk factor is coded from 0 to 2 with the same meanings associated with each rating. After rating the items, evaluators then make final risk ratings of low, moderate or high risk across each of the seven risk domains, including violence toward others. Similar to the HCR-20, evaluators are encouraged to consider the number of strength and vulnerability factors present, the combination of strength and vulnerability items, the relevance of each of these risk factors to the case at hand, and the types and intensity of services required to mitigate this person's need. For the current study, the total numbers of strength and vulnerability factors were totalled to arrive at two dynamic scales.

The START is a relatively new assessment instrument and has not been subject to as much empirical investigation as the HCR-20. Nevertheless, several examinations of the START have been conducted that support its reliability and validity for assessing risk for violence. That is, previous research has found the START to have good predictive validity and inter-rater reliability (Braithwaite, Charette, Crocker, & Reyes, 2010; Chu, Thomas, Ogloff, & Daffern, 2011; Desmarais, Nicholls, Wilson, & Brink, 2012; Gray et al., 2011; Nicholls, Brink, Desmarais, Webster, & Martin, 2006; Nonstad et al., 2010; Wilson et al., 2013). For instance, Nicholls and colleagues (2006) found excellent inter-rater reliability (ICC = .87) and fair to good predictive validity with regards to physical violence (AUC = .70). As well, Wilson et al. (2013) found inter-rater reliabilities for the strength and vulnerability total scores of .85 and .90, respectively. Moreover, the START was predictive of future violence (AUCs = .82 to .89) across a 12-month follow-up period.

## *Definition and Measures of Violence*

The definition of violence adopted in the current research was that provided by the HCR-20 (Webster et al., 1997). Violence is any "actual, attempted, or threatened harm to a person or persons" (p. 24). Violence was measured at each assessment period using two sources: a semi-structured interview with the participants and file information (correctional files were available for both samples, as well as community mental health files for the civil psychiatric sample only). In the interviews, violence is documented according to the MacArthur Community Violence Interview (Monahan et al., 2001). This system asks about nine type or categories of violence perpetration, that

range from throwing something at someone, to pushing, choking, or using a gun on someone. The final category consists of an "other" option for additional violent incidents that may meet the definition of violence but are not captured by the other eight categories of violence. As well, some additional categories of violence were assessed during the interview in order to more accurately reflect the definition of violence provided within the HCR-20 manual. These additional categories include acts such as threats of violence without any weapons in hand and deliberately causing fear in another person. The different categories of violence were collapsed in the present study into two dichotomous perpetration of violence outcomes. The overall, or broad, operationalization of violence included any of the above categories of violence, whereas the narrow definition of violence included only the more serious forms of violence (i.e., those categories included the MacArthur Community Violence Interview).

# Procedure

## *Study design and assessment schedule*

The current project employed a prospective, repeated measures longitudinal research design. Each participant was assessed at baseline and then once a month for up to five months. Therefore, each participant was assessed up to six times. Each assessment included a semi-structured interview and a review of the participants' files. The interview inquired about a broad range of information relevant to rating the HCR-20 and START, including information regarding mental health issues, substance use, social support networks, treatment and services, various attitudes and behaviour, and future plans. For the civil psychiatric sample, assessors reviewed their hospital records regarding the index hospitalization during the baseline phase. During the follow up phase, for those participants attending selected community mental health outpatient clinics, assessors reviewed these outpatient files. For the correctional sample, assessors accessed participants' correctional files during all phases of the study so long as the participant was still under the supervision of provincial corrections. The baseline interview took place during or shortly after the index hospitalization for the civil psychiatric sample and prior to release for the incarcerated correctional sample. The Clinical and Risk Management items on the HCR-20 and the entire START protocol

were completed at each of the six assessments. Violence was also assessed at each of the six assessments.

### *Assessors*

Assessors for the present study were all students enrolled in clinical or experimental psychology programs, varying in experience from Honours undergraduate students to senior level doctoral students. Assessors received specialized training for the current study in interviewing and completing the assessment protocols and instruments. In terms of interviewing, assessors were familiarized with the semi-structured interview, observed qualified individuals conduct the interview and were then observed over their first cases. Thus, it was assured that the assessors were properly trained and conducting the interview in a standardized manner. Assessors also received specialized training on the above-mentioned measures. An author of each of the relevant measures provided the training. Training included assigned readings, a review of the respective manuals, completion of video practice cases, and practice cases that were reviewed with the trainer. In addition, multiple assessors coded all initial cases until acceptable levels of interrater reliability were achieved.

## Statistical Analyses

For the analyses described below, the two subsamples of interest were combined in order to increase the number of data points included in each set of analyses. Due to the largely exploratory nature of the analyses, the unique characteristics of the risk assessment instruments within each subsample were not the focus of the present study. That is, as discussed in more detail below, the focus of the current study was not to generalize the findings to all civil psychiatric patients or correctional offenders. As well, the relatively small size of each of the subsamples, especially when only considering those cases with complete follow-up data, restricted the ability to investigate many of the research questions separately within the subsamples. As a result, the two subsamples were combined for the following analyses.

Missing data on the assessment instruments (e.g., the HCR-20 and START) was dealt with by prorating the scale scores according to instructions provided in the manuals. For the HCR-20, scores on the C and R scales were prorated if no more than one item was missing from each scale, respectively. Only one assessment contained an omitted item on the HCR-20 C scale. For the START, scores on the Strength and Vulnerability scales were prorated if no more than two items were missing from each scale, respectively. A total of 19 assessments were missing one item on the START Strength scale, while three assessments were missing two items. A total of 14 assessments were missing one item on the START Vulnerability scale, while three were missing two items. In order to prorate the scales, the average item score from the completed items was multiplied by the total number of items on the scale.

In order to address inter-rater reliability of the risk measures, two raters independently coded approximately 25% of the baseline assessments in the civil psychiatric sample only ($N$ = 31). Intraclass correlation coefficients (ICCs) were used to evaluate inter-rater reliability. Intraclass correlations provide a chance corrected measure of agreement rather than simply association (such as Pearson's r) and are mathematically equivalent to a weighted kappa (Bartko & Carpenter, 1976; Cicchetti & Sparrow, 1981; Douglas et al., 2003; Landis & Koch, 1977). A two-way random effects model evaluating absolute agreement was employed in the present study. Several authors have also called for the need to provide categorical descriptions that align with numerical reliability ranges. For instance, Cicchetti and Sparrow (1981), describing kappa and weighed kappa coefficients, defined reliability indices of below .40 as poor, .40 to .59 as fair, .60 to .74 as good, and above .75 as excellent. In comparison, Landis and Koch (1977) described indices of .00 to .20 as slight, .21 to .40 as fair, .41 to .60 as moderate, .61 to .80 as substantial, and above .81 as almost perfect.

## *Presence and rate of change: Change scores and reliable change index*

In order to address the first research question, the general question of whether dynamic risk factors do in fact change over time, change scores and a reliable change index (RCI) were calculated for each pair of assessments. Simple change scores (post-test minus pre-test) are provided for each pair of assessments. As well, RCIs allow for

the examination of within individual change over time by calculating a change score and comparing that change score with a critical value based on the upper 97.5% of the standard normal distribution (Draycott et al., 2012). In this sense, RCI controls for the possibility that any observed change is due to chance or measurement error and not actual change. The method employed in the current project is that described by Jacobsen and Truax (1991) in which the RCI is equivalent to the individual's post-treatment score minus their pre-treatment then divided by the standard error of the difference. The RCI, thus, provides a standardized change score for each assessment pair that can be compared to the critical value of 1.96, which equates to using an alpha level of 0.05 in traditional parametric null hypothesis statistical testing. Additional ways of calculating RCIs are available that take practice effects or regression to the mean into account. However, the present study employed the Jacobson and Truax method as practice effects are not relevant in clinician rated structured risk assessment instruments. Moreover, this procedure is favoured for several reasons including its computational simplicity and lack of substantial differences observed between the various methods in comparison studies (e.g., McGlinchey, Atkins, & Jacobson, 2002; McGlinchey & Jacobson, 1999; Speer & Greenbaum, 1995; Wise, 2004).

In order to answer the second research question regarding the rate of change in dynamic risk factors, change scores and RCIs were calculated within three different reassessment intervals: reassessments that took place in 30 days or less, between 31 and 59 days, or 60 days or more. In doing so, the amount of change across these categories can be observed for each of the dynamic scales. This allows for conclusions to be drawn about the rate of change of the different dynamic scales and the optimal length of the reassessment period, in terms of the assessment period that captures the most change, depending on the scale being used.

For these analyses, assessments from each follow-up were collapsed across participants in order to obtain a total number of assessment pairs across the entire sample. That is, the level of analyses for the RCI is that of the assessment pair. For these analyses, each assessment pair (e.g., baseline to follow-up 1, follow-up 1 to follow-up 2, etc.) was included as a separate case. As such, a given participant may contribute up to five observations to these analyses. Although assessment pairs are the unit of analysis, these analyses are still investigating intraindividual change from one

assessment to the next; however, each individual may contribute more than one observation to the total number of assessment pairs. This procedure essentially reduces the study design from a multiple time-point evaluation to a dual time-point evaluation. This analysis does not require independence of observations. Thus, the inclusion of up to five observations from a given participant does not violate any assumptions of the procedure.

When participants were missing data for entire assessment periods, these assessments were excluded from these analyses. Multiple imputation was not employed due to the exploratory nature of the analyses. As well, multiple imputation was deemed inappropriate, as this procedure may in fact lead to increased error and distort various statistical coefficients, especially when there is a high proportion of missing data, and when combined with statistical procedures for analyzing change over time multiple imputation may mask change in observed scores (e.g., Enders, 2011; Garson, 2012b; Grittner, Gmel, Ripatti, Bloomfield, & Wicki, 2011; Newman, 2003; Rubin, 1987).

### *Type and subtypes of change: ANOVA and Cluster analysis*

In order to answer the third research question regarding the type of change across individuals, a repeated measures ANOVA was employed to determine the overall type of change over time for each of the four dynamic scales for the entire sample. This procedure determines at the group level whether there is a change in the mean scores on each dynamic scale across the assessments. This analysis, although considered inappropriate for assessing whether in fact intraindividual change is occurring, allows for the determination of the overall type of change seen in the entire sample. That is, this procedure allows for the identification at the group level of the general pattern or shape of change seen in each of the dynamic scales.

Subsequently, in order to answer the fourth research question, cluster analysis was employed in order to find homogenous subgroups of individuals within the sample. Proceeding from the identification of the overall type or shape of change in the entire sample, cluster analysis was used to determine if certain groups of individuals showed particular trajectories of change over time. Cluster analysis is an exploratory

classification procedure that assigns individuals to certain groups, or clusters, based on their similarity on a number of variables (Antonenko, Toy, & Niederhauser, 2012; Garson, 2012a; Pell & Hargreaves, 2011). Cluster analysis attempts to do this by both minimizing within-group variation and maximizing between-group variation (Douglas et al., 2011; Garson, 2012a). Each of the participants' scores on each of the dynamic scales were entered as clustering variables into separate hierarchical clustering analyses, for each scale respectively, in order to identify the number of clusters, or the number of groups, that change in different ways. In order to determine the optimum number of clusters, hierarchical cluster analyses was used as this procedure begins with each individual as their own cluster and then begins to group individuals together based on the similarities and differences between their scores on the risk assessment measures until only one cluster remains (Garson, 2012a). Clusters were formed according to the distance between participant's scores on the respective risk instruments using squared Euclidean distances, which establish the distance, or how far apart, two cases were across all of the assessments. As well, the Ward's linkage method was employed which calculates the sum of the squared Euclidean distances from every case in a given cluster to the mean of all the variables and then merges clusters in order to minimize this sum, similar to an analysis of variance approach. This method, thus, is meant to maximize between-group differences and minimize within-group differences.

This exploratory procedure, as mentioned, begins by placing each individual in his or her own cluster and then merging clusters until only one remains (Antonenko et al., 2012; Garson, 2012a). Following this, the researcher must choose the optimal number of clusters based on the squared Euclidean distances. This was accomplished in the present research by plotting the agglomeration coefficients (the sum of the squared Euclidean distances at each stage) in a scree plot and identifying a distinctive change, or break, in the slope of the line. This change in slope corresponds to an identifiable jump in the increase of coefficients from one stage to the next.

These analyses only included those participants that included data for each of the assessment times on a given measure (i.e., listwise deletion). That is, only those participants that were assessed all six times on a given dynamic risk scale were included in these analyses (see Table 1). Once again, multiple imputation was not undertaken due to the exploratory nature of the analyses. The exact manner in which groups of

individuals may change in different ways is of less concern than the ability to identify such groups. Generalization of the specific manner in which different groups changed in the present study is not the intention of these analyses, as numerous sample and individual characteristics will likely impact a given individual's, or group's, trajectory of change. Moreover, as discussed above, multiple imputation is not without drawbacks, including the possibility of increasing error and distorting statistical coefficients, as well as the possibility of concealing change in the observed scores (e.g., Enders, 2011; Garson, 2012b; Grittner et al., 2011; Rubin, 1987; Newman, 2003).

### *Association of change with violence: Generalized estimating equations*

In order to answer the final research question regarding the association of change in dynamic risk factors with subsequent violence, the method of generalized estimating equations (GEE) was employed (Liang & Zeger, 1986; see also Diddle, Liang, & Zeger, 1994). Specifically, GEE was used to determine whether change on the dynamic risk scales was statistically associated with violence in the following assessment interval. GEE is an extension of the generalized linear model and is ideally suited for repeated measures designs and non-linear functions (Garson, 2013; Liang & Zeger, 1986). GEE also does not hold many of the assumptions of other generalized linear models. As such, GEE can handle multiple dichotomous outcomes and correlated repeated measures within subjects. GEE was used to estimate the independent statistical association between change on the dynamic risk scales and violence over the six assessments providing odds ratios for each of the dynamic scales (Michel et al., 2013). Separate models were run for each of the four dynamic risk scales over the six assessments. The current model employed a binominal distribution with a binary logistic link function, as the dependent variable (i.e., violence) was dichotomous. Additionally, a robust error estimator method was employed which provides better estimates of error terms for large datasets and the independent correlation structure was employed for the correlation matrix. All available observed data was entered into this set of analyses. Multiple imputation was not employed, as this procedure accommodates different numbers of observations per case.

# Results

## Sample Size and Attrition

As seen in Table 1, a total of 216 participants were assessed on the dynamic risk scales at the baseline phase, followed by 161, 134, 119, 110, and 100 participants at each of the follow-ups, respectively. This corresponds to 74.5% of the sample being assessed at least twice and 46.3% of the sample being assessed the full six times. Table 1 also presents the mean duration of each assessment interval. On average the assessments were approximately one-and-a-half months apart.

***Table 1. Sample Size and Duration of Assessment Interval across Assessments***

| Assessment | Attrition | | Duration |
|---|---|---|---|
| | N | % Total | Days |
| Baseline | 216 | 100% | - |
| Follow-up #1 | 161 | 74.5% | 53.43 (35.65) |
| Follow-up #2 | 134 | 62.0% | 46.17 (25.23) |
| Follow-up #3 | 119 | 55.1% | 45.39 (28.29) |
| Follow-up #4 | 110 | 50.9% | 42.91 (20.10) |
| Follow-up #5 | 100 | 46.3% | 38.23 (12.25) |

*Note.* Values provided are means followed by standard deviations.

Several comparisons were made between those who completed at least one follow-up and those who did not complete any follow-ups. Compared to participants that did not complete any follow-ups, participants who completed at least one follow-up did not differ based on their age, $t(233) = -0.93$, $p = .355$, gender, $\chi^2(1, N = 235) = 1.33$, $p = .249$, or ethnicity, $\chi^2(5, N = 233) = 3.89$, $p = .566$. No differences were found between these groups based on their baseline scores on the HCR-20 H scale, $t(214) = 0.38$, $p = .704$, C scale, $t(214) = 0.26$, $p = .798$, or R scale, $t(214) = -0.10$, $p = .918$, as well as the

START Vulnerability scale, $t(213) = 0.45$, $p = .653$, or START Strength scale, $t(213) = -0.98$, $p = .328$. Finally, no differences were found in baseline PCL scores $t(212) = -0.29$, $p = .773$, rates of previous serious violence, $\chi^2(1, N = 214) = 0.66$, $p = .416$, or rates of previous overall violence $\chi^2(1, N = 214) = 1.19$, $p = .276$.

Comparisons were also made between those who completed all six assessments and those who did not complete all six assessments. Compared to participants that did not complete all six assessments, participants who completed all of the assessments did not differ based on their age, $t(233) = -1.56$, $p = .119$, gender, $\chi^2(1, N = 235) = 0.16$, $p = .694$, or ethnicity, $\chi^2(5, N = 233) = 3.70$, $p = .593$. As well, no differences were found between these groups on their baseline scores on the HCR-20 H scale, $t(214) = 1.79$, $p = .074$, or R scale $t(214) = 1.39$, $p = .166$, as well as their baseline rates of previous serious violence, $\chi^2(1, N = 214) = 0.07$, $p = .786$, or rates of previous overall violence $\chi^2(1, N = 214) = 0.00$, $p = .959$. However, participants who completed all six assessments were found to have lower baseline scores on the HCR-20 C scale, $t(214) = 2.14$, $p = .033$, $d = 0.29$, 95% CI [-0.01, 0.60], START Vulnerability scale, $t(213) = 2.44$, $p = .016$, $d = 0.34$, 95% CI [-0.64, 1.31], and PCL, $t(212) = 2.58$, $p = .011$, $d = 0.35$, 95% CI [-0.38, 1.08], as well as higher scores on the START Strength scale, $t(213) = -2.86$, $p = .005$, $d = 0.39$, 95% CI [-0.76, 1.55].

## Descriptive Statistics of Dynamic Risk Assessment Scales and Violence

Table 2 presents the percentage of the sample that perpetrated violence during each interval and the means and standard deviations of the four dynamic scales at each assessment point. A substantial proportion of the sample had committed at least one act of violence in the six months prior to the baseline interview (33.2%) with many of these being more serious incidents of violence (29.9%). A smaller proportion of the sample perpetrated violence during each of the subsequent assessment intervals, as seen in Table 2.

***Table 2. Descriptive Statistics of Risk Measures and Violence across Assessments***

| Assessment | Violence | | HCR-20 | | START | |
|---|---|---|---|---|---|---|
| | Broad | Narrow | C Scale | R Scale | Vulnerability | Strength |
| Baseline | 33.2% | 29.9% | 4.44 (2.31) | 5.08 (2.40) | 17.85 (7.44) | 19.69 (8.81) |
| Follow-up #1 | 16.1% | 9.3% | 3.56 (2.30) | 4.50 (2.58) | 15.13 (8.17) | 22.14 (8.81) |
| Follow-up #2 | 18.7% | 11.2% | 3.72 (2.64) | 4.78 (2.66) | 15.21 (9.98) | 22.47 (9.12) |
| Follow-up #3 | 10.8% | 8.3% | 3.21 (2.14) | 4.17 (2.51) | 14.05 (8.22) | 23.87 (9.12) |
| Follow-up #4 | 12.7% | 7.3% | 3.20 (2.38) | 4.15 (2.63) | 13.98 (8.57) | 24.18 (9.69) |
| Follow-up #5 | 11.0% | 6.1% | 3.04 (2.35) | 3.85 (2.37) | 13.34 (7.92) | 23.94 (9.27) |

*Note.* Values provided are means followed by standard deviations.

## Inter-rater Reliability

The ICCs and categorical descriptors of the inter-rater reliability of the four dynamic scales based on a subsample of 31 civil psychiatric inpatients are listed in Table 3. Reliability coefficients ranged from .42 to .76 for the dynamic scales. The START Vulnerability scale produced the largest coefficient falling in the excellent or substantial range, followed by the HCR-20 C scale in the good or substantial range, then the START Strength scale and HCR-20 R scale both in the fair or moderate range.

***Table 3. Inter-rater Reliability of Risk Measures***

| Dynamic Scale | Numerical Values | | Categorical Descriptors | |
|---|---|---|---|---|
| | $ICC_1$ | (95% CI) | Cicchetti & Sparrow, 1981 | Landis & Koch, 1977 |
| HCR-20 C Scale | .63 | (.36 - .80) | Good | Substantial |
| HCR-20 R Scale | .42 | (.09 - .66) | Fair | Moderate |
| START Vulnerability | .76 | (.56 - .88) | Excellent | Substantial |
| START Strength | .52 | (.21 - .74) | Fair | Moderate |

## Research Question 1: Change in Dynamic Risk Scales Over Time

Change scores and RCIs were calculated to determine whether intraindividual change is seen on the four dynamic scales from one assessment to the next. Table 4 presents the descriptive statistics (means and standard deviations) of the scales when the data was rearranged so that each assessment pair was the unit of analysis. When collapsed in this manner, a total of 622 assessment pairs were included in the analyses.

***Table 4. Descriptive Statistics of Risk Measures for RCI Analyses***

| Assessment | Duration | HCR-20 | | START | |
|---|---|---|---|---|---|
| | Days | C Scale | R Scale | Vulnerability | Strength |
| Pre | 45.75 (26.02) | 3.65 (2.39) | 4.53 (2.53) | 15.14 (8.18) | 22.39 (9.03) |
| Post | | 3.39 (2.38) | 4.32 (2.56) | 14.37 (8.31) | 23.05 (9.01) |

*Note.* Values provided are means followed by standard deviations.

Figure 1 presents the frequencies of raw change scores, in percentages of the total number of assessment pairs, for the HCR-20 C and R scales for all assessments pairs. As seen in Figure 1, many of the scale scores did in fact change from one assessment to the next, with the majority of scores changing by only one or two points in either direction. However, some scale scores changed by as much as six to nine points; the latter of which corresponds to change on every single item in the scale.

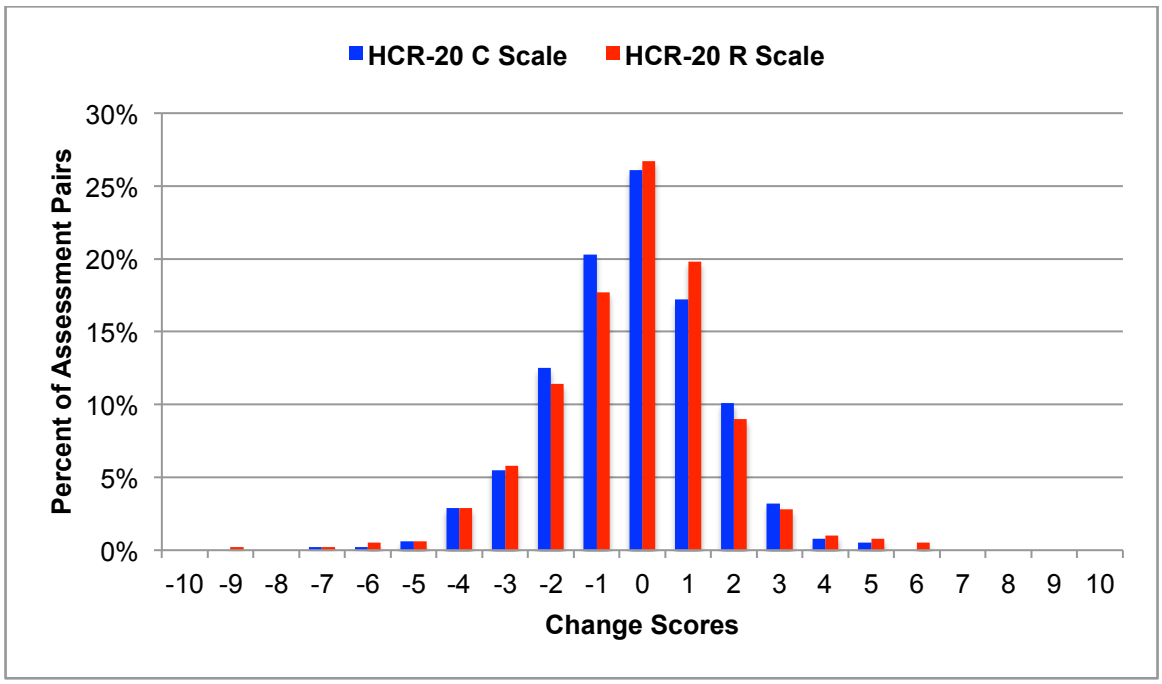


***Figure 1. HCR-20 Clinical and Risk Management Scale Change Scores***

Figure 2 presents the same information for the START Strength and Vulnerability scales across the assessment pairs. Once again, change was seen in a substantial proportion of the assessment pairs. The majority of scores changed by one to five points; however, some scores changed by as much as 18 to 27 points, the latter of which corresponds to change on at least 13 items.

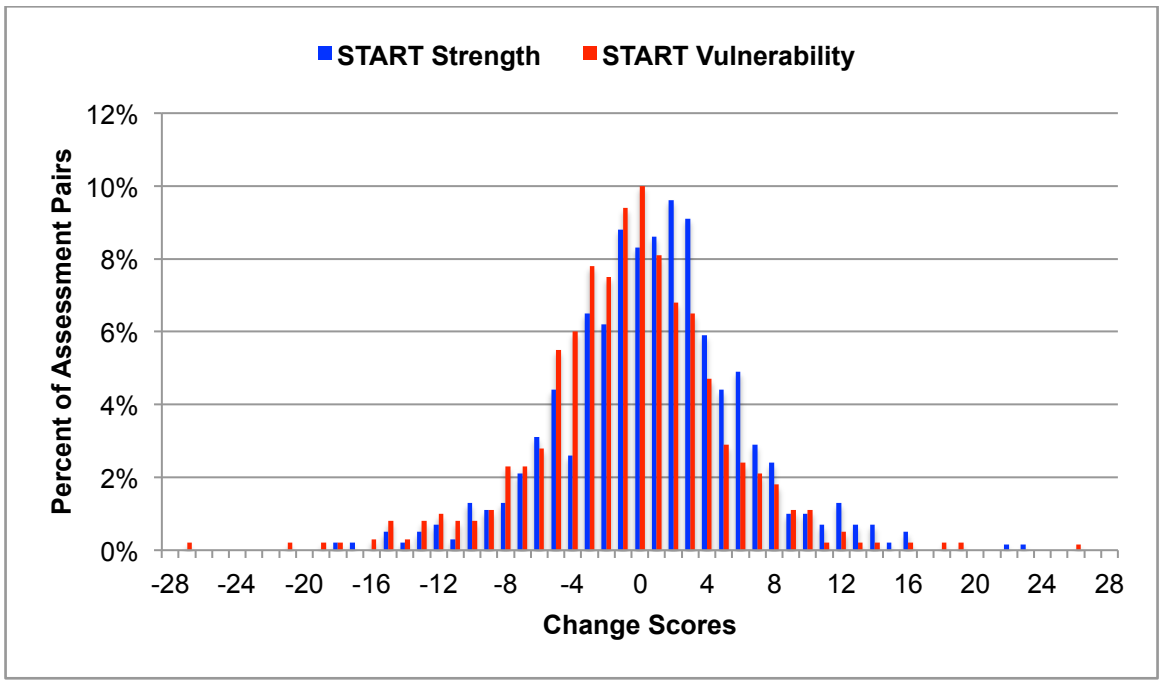


***Figure 2. START Strength and Vulnerability Scale Change Scores***

Table 5 presents the RCIs for the four dynamic scales for the entire sample (as well as separated by the length of the assessment interval which is discussed below). Based on Figures 1 and 2 and Table 5, it is apparent that many of the scores on the four dynamic scales did change from one assessment to the next. Only 26.7% of the scores did not change on the HCR-20 R scale from one assessment to the next, followed by 26.0% on the HCR-20 C scale, 10.0% on the START Vulnerability scale and 8.3% on the START Strength scale. However, much of the observed change did not reach the level of reliable change according to the criteria used for these analyses. Reliable change was only seen, in descending order, in 6.6% of the HCR-20 R scale scores, 6.6% of the START Strength scores, 6.3% of the START Vulnerability scores, and 5.2% of the HCR-20 C scale scores. Thus, intraindividual change was seen in a large proportion of the assessments, but this change rarely met the threshold for reliable change.

*Table 5. Reliable Change Indexes of the HCR-20 and START Assessments*

| | HCR-20 | | START | |
|---|---|---|---|---|
| | C Scale | R Scale | Vulnerability | Strength |
| **Overall Sample (N = 622)** | | | | |
| Reliable Increase | 1.3% | 2.3% | 1.6% | 4.2% |
| Increase | 30.5% | 31.7% | 37.6% | 47.6% |
| No Change | 26.0% | 26.7% | 10.0% | 8.3% |
| Decrease | 38.2% | 35.0% | 46.0% | 37.5% |
| Reliable Decrease | 3.9% | 4.3% | 4.7% | 2.4% |
| **30 Days or Less (N = 107)** | | | | |
| Reliable Increase | 0.9% | 1.9% | 1.9% | 1.9% |
| Increase | 24.3% | 32.7% | 38.1% | 43.8% |
| No Change | 28.0% | 31.8% | 12.4% | 8.6% |
| Decrease | 44.8% | 29.9% | 44.7% | 42.9% |
| Reliable Decrease | 1.9% | 3.7% | 2.9% | 2.9% |
| **31 to 59 Days (N = 420)** | | | | |
| Reliable Increase | 1.0% | 1.7% | 1.2% | 4.6% |
| Increase | 30.4% | 30.0% | 36.1% | 49.4% |
| No Change | 26.9% | 27.0% | 9.8% | 8.7% |
| Decrease | 37.4% | 37.5% | 47.9% | 34.7% |
| Reliable Decrease | 4.3% | 3.8% | 5.0% | 2.7% |
| **60 Days or More (N = 95)** | | | | |
| Reliable Increase | 3.2% | 2.1% | 2.1% | 5.3% |
| Increase | 37.9% | 41.1% | 44.7% | 43.6% |
| No Change | 20.0% | 20.0% | 8.5% | 6.4% |
| Decrease | 34.7% | 32.6% | 41.5% | 43.6% |
| Reliable Decrease | 4.2% | 4.2% | 3.2% | 1.1% |

Several patterns are apparent in Table 5. It appears as though scores were slightly more likely to decrease across the assessments on the HCR-20 C (42.1% decrease versus 31.8% increase) and R scales (39.2% versus 34.0%), as well as the START Vulnerability scale (50.7% versus 39.2%), whereas scores were more likely to

increase on the START Strength scale (39.9% versus 51.8%). Additionally, it appears that scores were slightly more likely to change on the START than on the HCR-20 from one assessment to the next, as the percentages of scores that changed and changed reliably are consistently higher for the START compared to the scale scores on the HCR-20.

## Research Question 2: Rate of Change in Dynamic Risk Scales Over Time

In order to address the second research question, inquiring whether the reassessment length affects change on the scales, change scores and RCIs were calculated on the assessment pairs separated into three categorical lengths. When separated by reassessment interval length into three categories (30 days or less, 31 to 59 days, and 60 days or more) the mean follow-up length in each category were 27.53 days (*SD* = 2.69), 39.62 days (*SD* = 7.33), and 93.38 days (*SD* = 37.24), respectively. Figures 3 and 4 present the frequencies of raw change scores, in percentages of the number of assessment pairs separated by assessment length, for the HCR-20 C and R scales, respectively. Figures 5 and 6 present the same information for the START Vulnerability and Strength scales, respectively. Table 5 presents the RCIs based on these reassessment lengths. Once again, when analyzed in these categories, a substantial amount of intraindividual change was seen in the scores from one assessment to the next on all four scales, yet this change often did not reach the level of reliable change according to these analyses

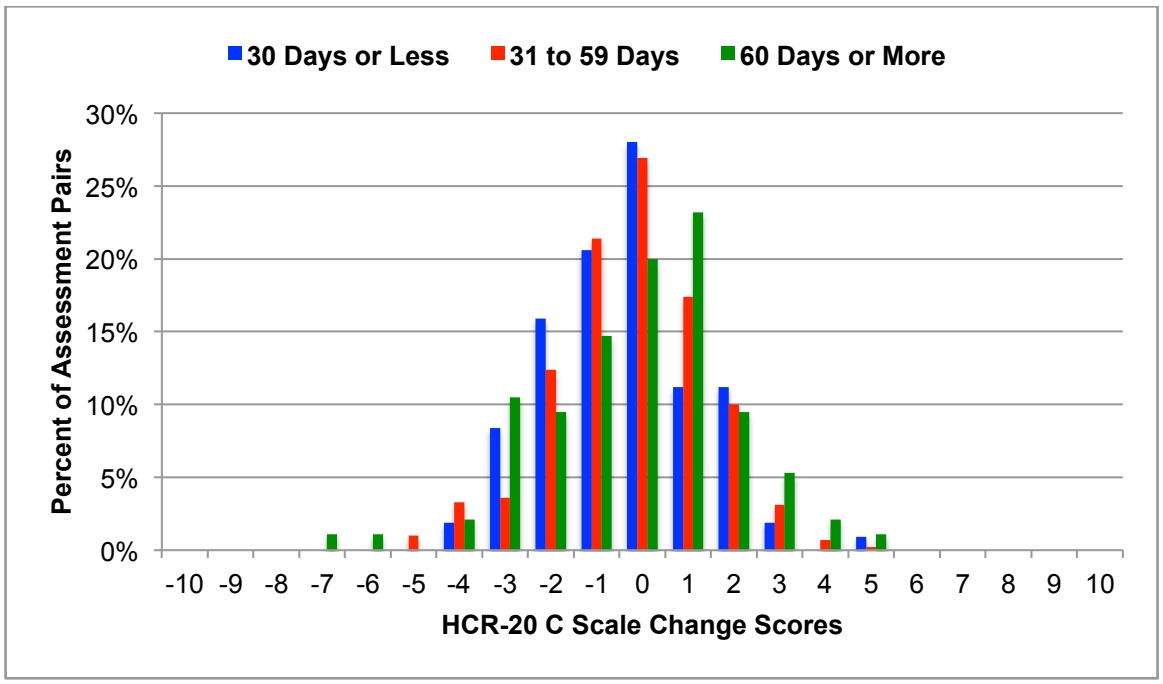


*Figure 3.* ***HCR-20 Clinical Scale Change Scores Separated by Reassessment Interval Length***

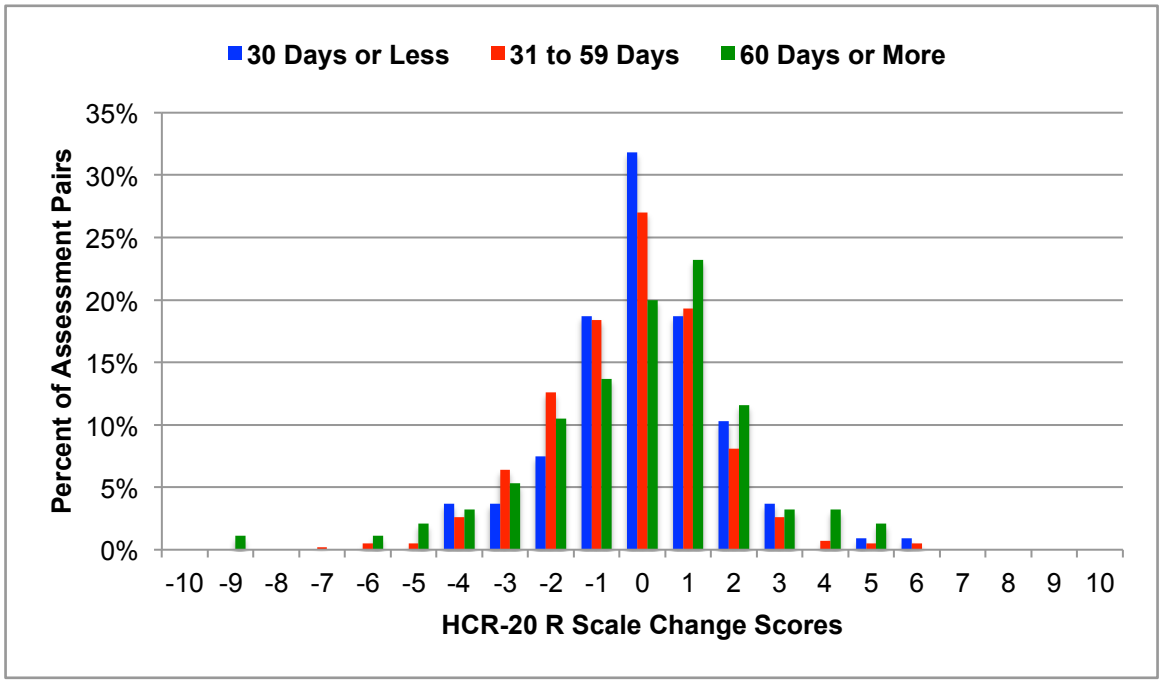


*Figure 4.* ***HCR-20 Risk Management Scale Change Scores Separated by Reassessment Interval Length***

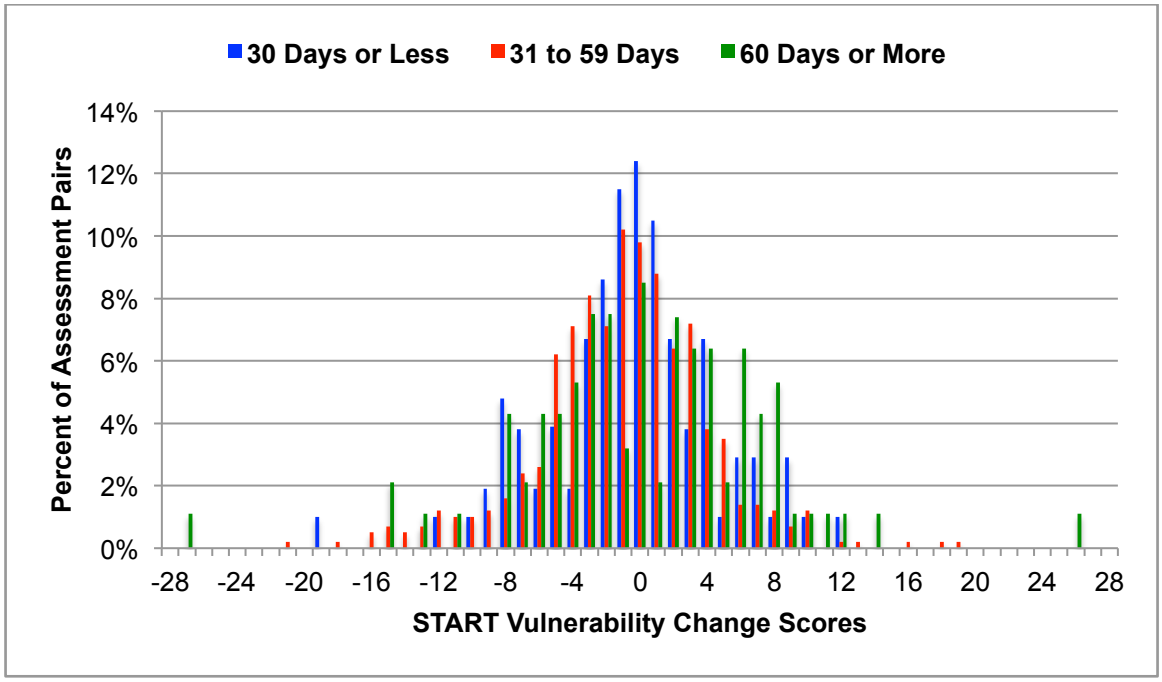


*Figure 5.* ***START Vulnerability Scale Change Scores Separated by Reassessment Interval Length***

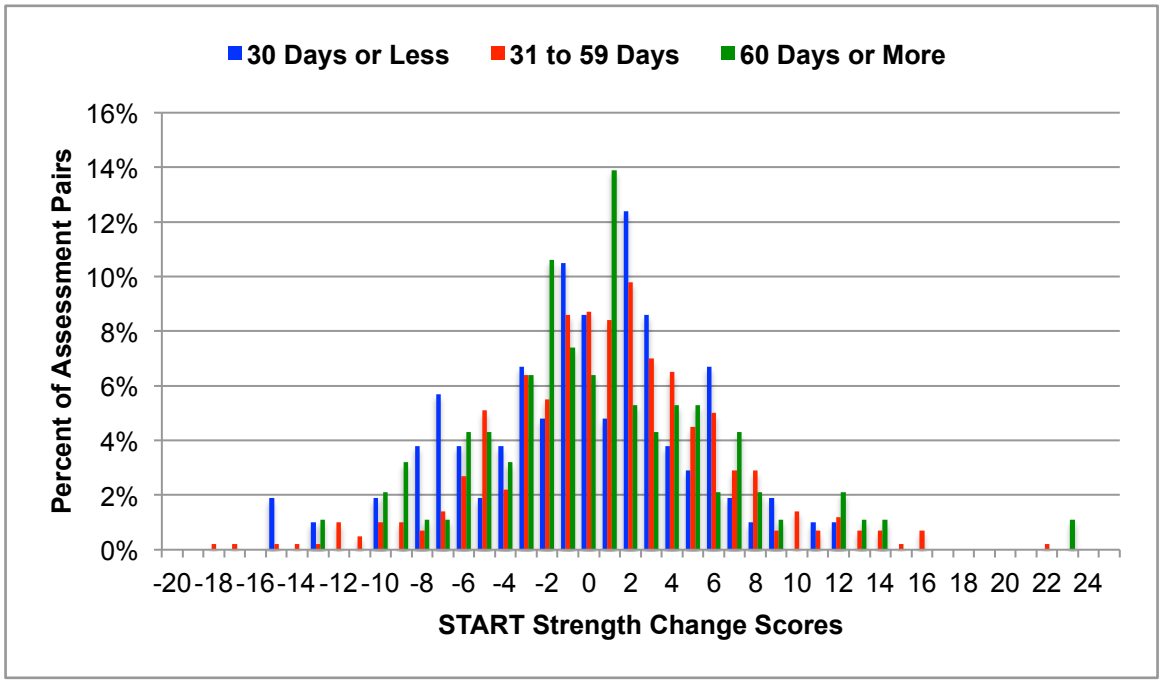


*Figure 6.* ***START Strength Scale Change Scores Separated by Reassessment Interval Length***

Several noteworthy patterns are evident in these results. Similar trends as reported above were seen in these analyses. When the assessment interval was 59 days or less, in general scores on the HCR-20 C and R scales, as well as the START Vulnerability scale tended to decrease more often than increase (average decrease of 44.2% versus 28.3% increase, 37.5% versus 33.1%, and 50.3% versus 38.7%, respectively), whereas scores on the START Strength scale more often increased (41.5% versus 49.9%). However, when the assessment interval was 60 days or more (green bars in the figures), all of the scales tended to increase more often than decrease (39.9% decrease versus 41.1% increase for the HCR-20 C scale, 36.8% versus 43.2% for the HCR-20 R scale, 44.7% versus 48.9% for the START Strength score, and 44.7% versus 46.9% for the START Vulnerability score). Once again, the START scales were slightly more likely to change compared to the HCR-20 scale scores, as the percentages of scores that changed and changed reliably are nearly consistently higher for the START than the HCR-20 in each of the different timeframes.

Additionally, it appears from the figures and table that for the HCR-20 scales less change was seen when the reassessment interval was shorter compared to a longer reassessment interval. For instance, with regards to the HCR-20 R scale, no change was seen in 31.8% of assessments done within 30 days or less, compared to 27.0% of assessments done in 31 to 59 days, and 20.0% of assessments done over 60 days or more. A similar pattern was seen for the HCR-20 C scale (28.0%, 26.9%, and 20.0% no change in each of the timeframes, respectively). The amount of assessments reaching the reliable change level also tended to increase as the length of the reassessment period increased. This pattern can also be seen in Figures 3 and 4; the distributions of change scores tend to be wider as the assessment length increases (i.e., the blue bars cluster closer to zero, the red bars extend farther towards the extremes, and the green bars tend to extend the farthest towards the extremes). However, for the START scales, the same patterns were not present. The START appeared to capture a more similar amount of change regardless of the length of the reassessment interval.

## Research Question 3: Type of Change in Dynamic Risk Scales Over Time

In order to investigate the overall type, or shape, of change at the sample level across all six assessments for the four dynamic scales, repeated-measures ANOVAs were employed. With regards to the HCR-20 C scale, the assumption of sphericity was violated, $\chi^2(14, N = 99) = 72.78$, $p < .001$. As a result, the Huynh-Feldt correction was employed, $\varepsilon = .76$. The mean scores were found to vary over the six assessments, $F(3.95, 386.93) = 9.34$, $p < .001$. Further analyses revealed that both a linear trend, $F(1, 98) = 17.32$, $p < .001$, $eta^2 = .15$, and quadratic trend, $F(1, 98) = 10.92$, $p = .001$, $eta^2 = .10$, fit the data. As well, for the HCR-20 R scale, the assumption of sphericity was also violated, $\chi^2(14, N = 99) = 45.77$, $p < .001$. As a result, the Huynh-Feldt correction was employed, $\varepsilon = .84$. Once again, the mean scores were found to vary over the six assessments, $F(4.38, 429.52) = 7.01$, $p < .001$. Further analyses revealed that the HCR-20 R scale tended to display a linear decrease over the six assessments, $F(1, 98) = 16.49$, $p < .001$, $eta^2 = .14$. Figure 7 presents the means of the HCR-20 scales across the six assessments. The mean decrease in scores across the six assessments is apparent in this figure, with some variability. Figure 7 also plots the percent of the sample that perpetrated violence at each assessment, which tends to follow a very similar pattern as both of the mean HCR-20 dynamic scale scores.

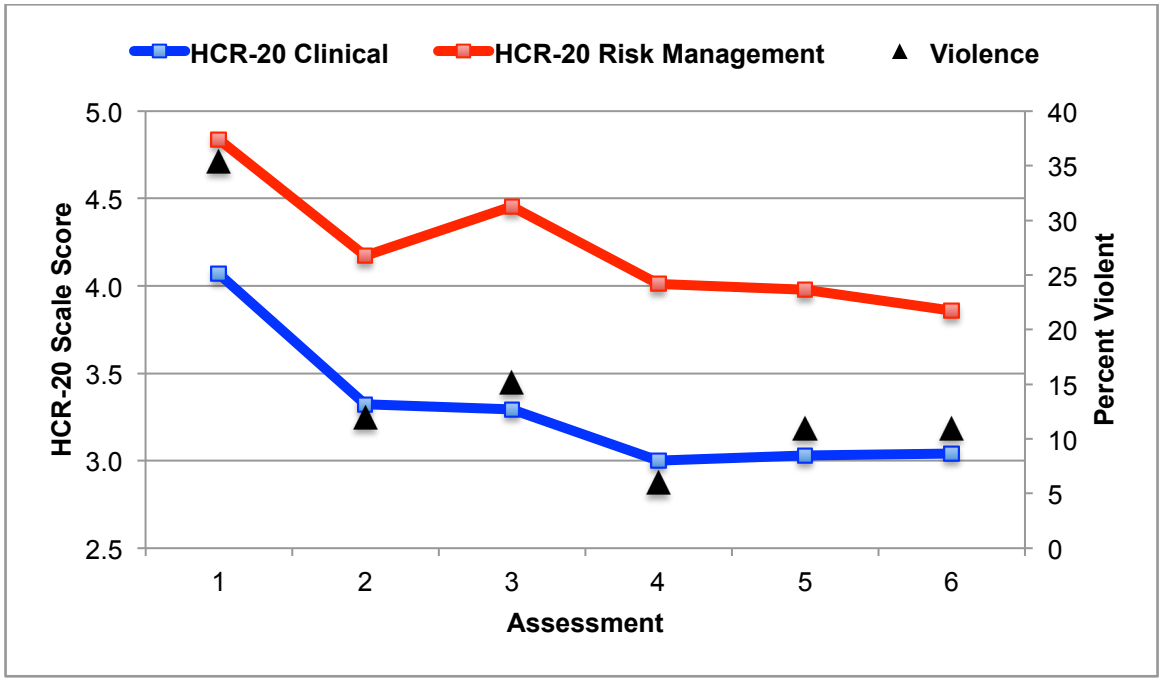


***Figure 7.*** ***Mean HCR-20 Clinical and Risk Management Scale Scores and Violence across Six Assessments***

With regards to the START Vulnerability scale, the assumption of sphericity was violated, $\chi^2(14, N = 97) = 26.69$, $p < .021$, so the Huynh-Feldt correction was employed, $\varepsilon = .89$. The mean scores varied over the six assessments, $F(4.71, 452.23) = 6.10$, $p < .001$. Both a linear trend, $F(1, 96) = 14.12$, $p < .001$, $eta^2 = .13$, and quadratic trend, $F(1, 96) = 6.64$, $p = .012$, $eta^2 = .07$, were found to fit the pattern of change. Finally, with regards to the START Strength scale, the assumption of sphericity was violated, $\chi^2(14, N = 97) = 43.65$, $p < .001$, and the Huynh-Feldt correction was employed, $\varepsilon = .86$. Once again, the mean scores were found to vary over the six assessments, $F(4.51, 432.73) = 8.723$, $p < .001$. Further analyses revealed that a linear trend, $F(1, 96) = 17.20$, $p < .001$, $eta^2 = .15$, a quadratic trend, $F(1, 96) = 12.23$, $p = .001$, $eta^2 = .11$, and a cubic trend, $F(1, 96) = 6.19$, $p = .015$, $eta^2 = .06$, fit the data. Figure 8 presents the means of the START scales across the six assessments, revealing the overall mean increase in START Strength and decrease in START Vulnerability scores. Figure 8 also plots the percent of the sample that perpetrated violence at each assessment. The mean START Strength scores appear to follow a similar, though inversed, trajectory as the perpetration of violence. However, little similarity is apparent between the mean START Vulnerability scores and the rate of violence perpetration at each of the assessments.


***Figure 8.*** ***Mean START Vulnerability and Strength Scale Scores and Violence across Six Assessments***

## Research Question 4: Group Differences in Trajectories of Change Over Time

Next, to address the question of different trajectories of change for different subgroups on the dynamic scales over the six assessments, hierarchical cluster analyses were used entering the scores on each of the dynamic scales, separately, over the six assessments as the clustering variables. With regards to the HCR-20 C scale, based on analysis of the scree plot presented in Figure 9, four clusters appear to fit the data best, which corresponds to a change in coefficients from 91.96 to 140.39. Figure 10 presents the mean scores for each of the four clusters across the assessments. As seen in Figure 10, one group of participants (Cluster A, *n* = 34) starts with low scores and then decreases gradually over time. Another group (Cluster B, *n* = 32) remains relatively constant with relatively low scores across the assessments. Another group (Cluster C, *n* = 16) starts relatively high, and then drops considerably before levelling off after the third assessment. The final group (Cluster D, *n* = 17) has fairly consistent high scores across the assessments with little change occurring.

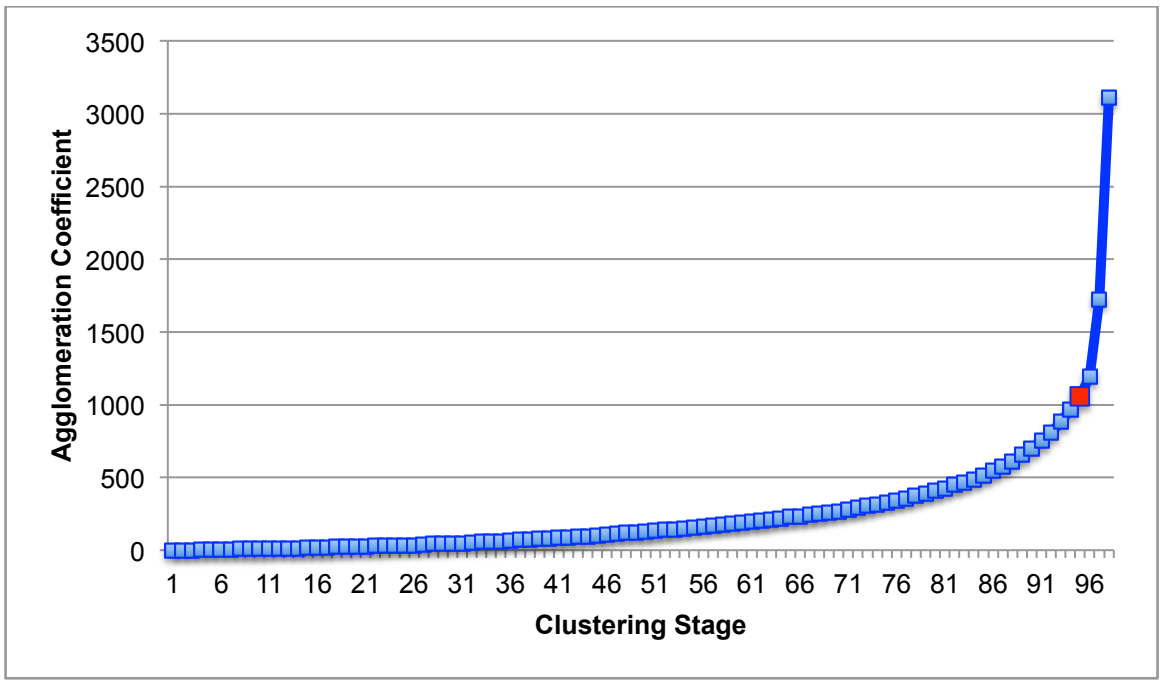


*Figure 9.* *Scree Plot of Agglomeration Coefficients for the HCR-20 Clinical Scale Scores*

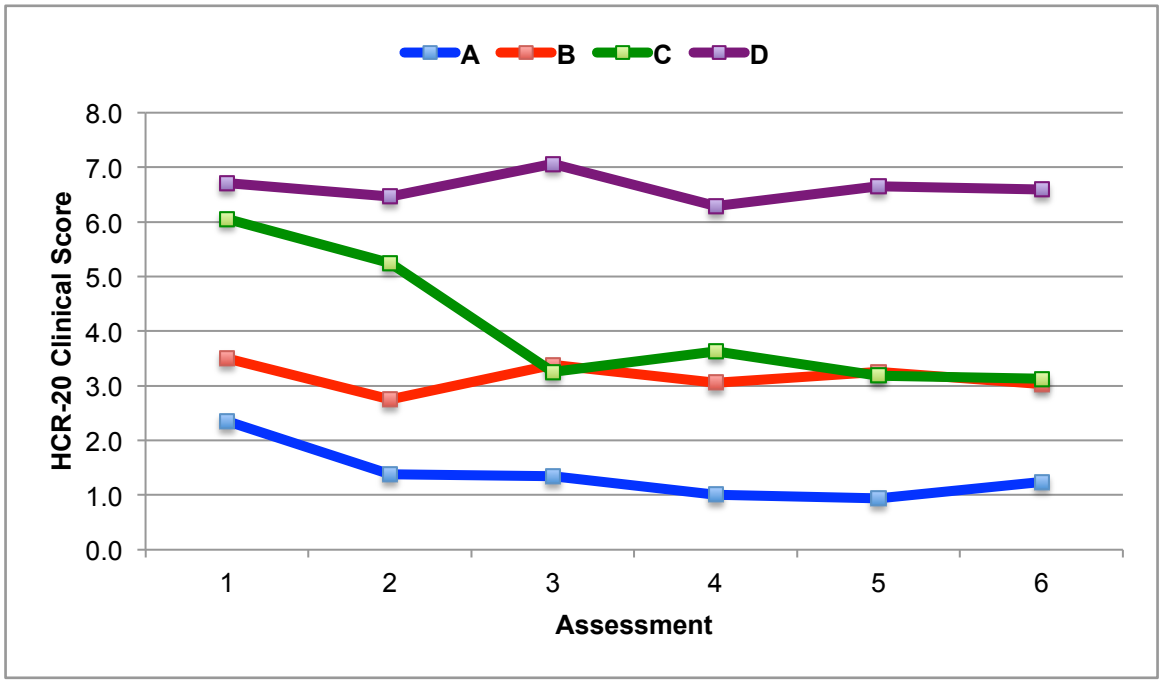


*Figure 10.* *Mean HCR-20 Clinical Scale Scores across Six Assessment Separated by Cluster*

For the HCR-20 R scale, based on analysis of the scree plot presented in Figure 11, four clusters also appear to fit the data best (change in coefficients from 109.66 to 247.15). Figure 12 presents the mean scores for each of the four clusters across the assessments. Once again, one group (Cluster A, *n* = 30) starts with low scores and then shows a slight, gradual decrease. A second group (Cluster B, *n* = 33) begins with scores in the middle range and then shows a steady decrease over the assessments. A third group (Cluster C, *n* = 25) starts with relatively higher scores and then increases before decreasing back to their starting scores. The fourth group (Cluster D, *n* = 11) has fairly consistent high scores with little change occurring.

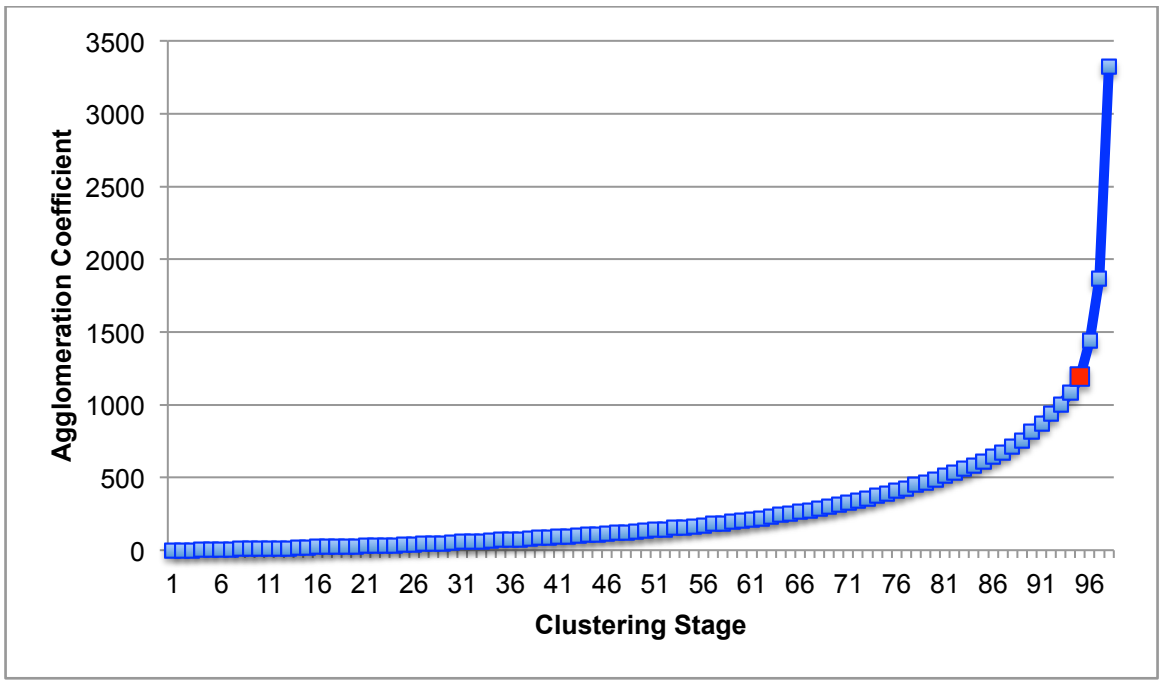


*Figure 11.* ***Scree Plot of Agglomeration Coefficients for the HCR-20 Risk Management Scale Scores***

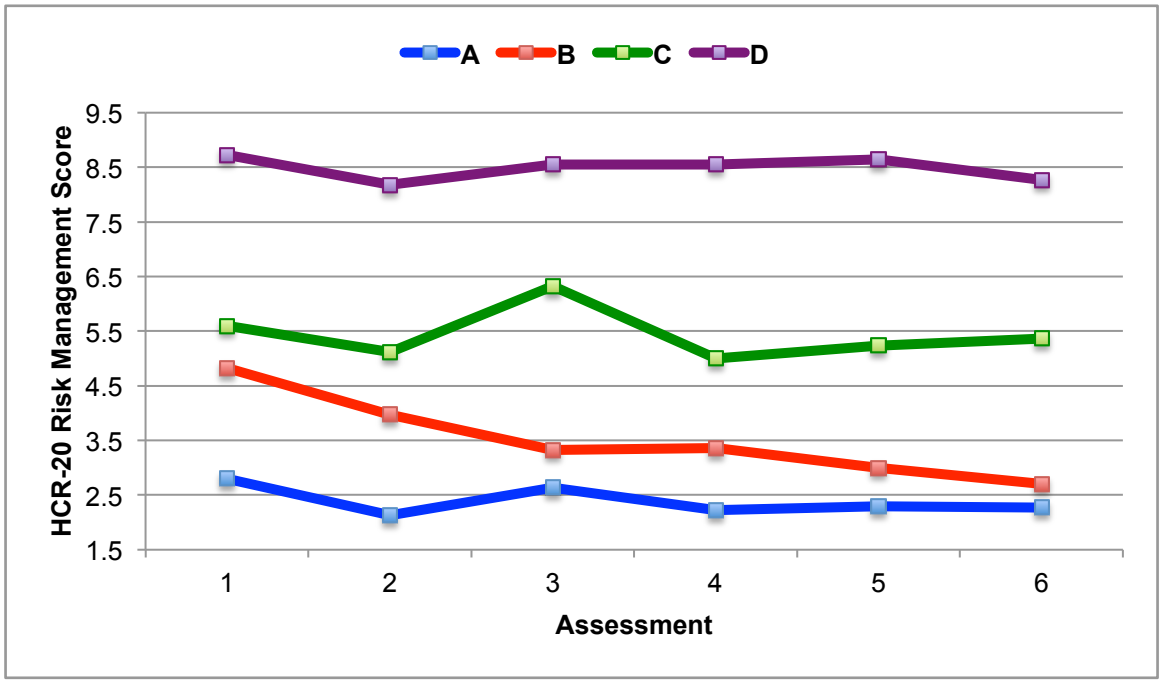


*Figure 12.* ***Mean HCR-20 Risk Management Scale Scores across Six Assessment Separated by Cluster***

Furthermore, with regards to the START Vulnerability scale, based on analysis of the scree plot presented in Figure 13, four clusters also appear to fit the data best (change in coefficients from 1294.59 to 1616.73). Figure 14 presents the mean scores for each of the four clusters across the assessments. One group (Cluster A, *n* = 43) starts with low scores and then decreases gradually in a fairly linear fashion over the assessments. Another group (Cluster B, *n* = 42) starts with scores in the middle range and then decreases from the first to second assessment before levelling off for the remainder of the assessments. Another group (Cluster C, *n* = 6) starts with scores in the middle range then waxes and wanes before increasing for the last two assessments. The final group (Cluster D, *n* = 6) starts with high scores and then increases slightly before decreasing slightly, thus, remaining with high scores.

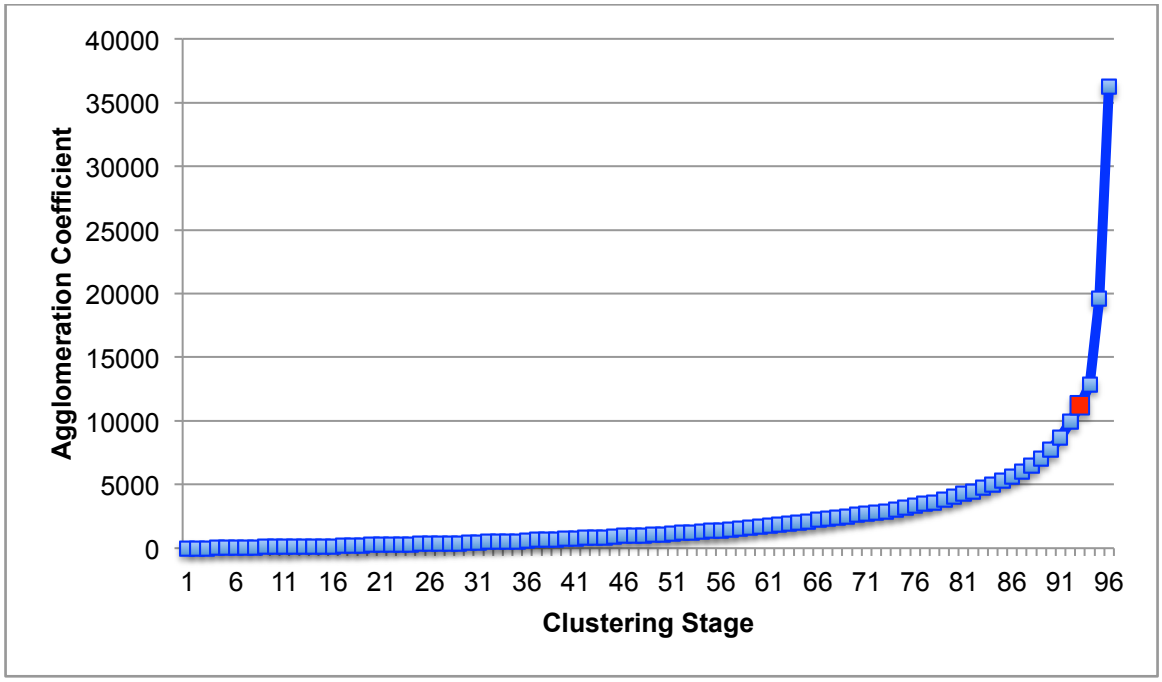


*Figure 13.* ***Scree Plot of Agglomeration Coefficients for the START Vulnerability Scale Scores***

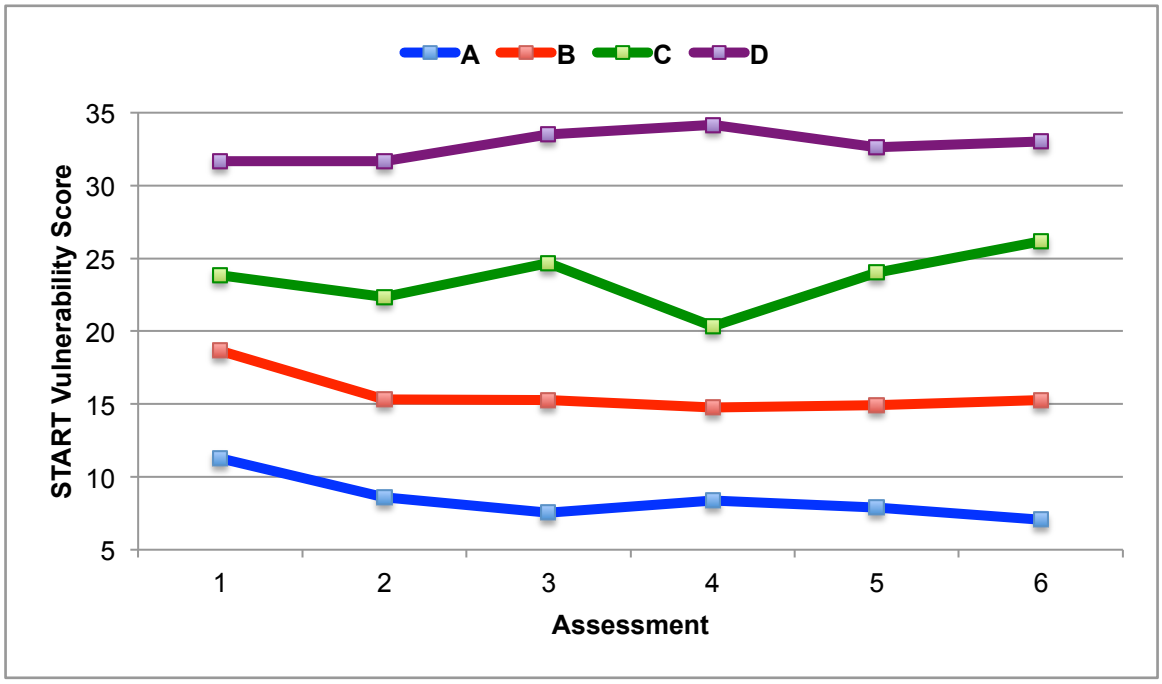


*Figure 14.* ***Mean START Vulnerability Scale Scores across Six Assessment Separated by Cluster***

For the START Strength scale, based on analysis of the scree plot presented in Figure 15, four clusters also appear to fit the data best (change in coefficients from 887.33 to 2783.20). Figure 16 presents the mean scores for each of the four clusters across the assessments. As seen with the other scales, one group (Cluster A, *n* = 26) starts with high scores and gradually increases slightly. A second group (Cluster B, *n* = 32) begins with scores in the middle range and then shows a steady increase over time until a drop at the final assessment. A third group (Cluster C, *n* = 23) starts scores in the middle range and then shows a slight increase before levelling off. The fourth group (Cluster D, *n* = 16) has fairly consistently low scores with relatively little change occurring, other than some slight waxing and waning.

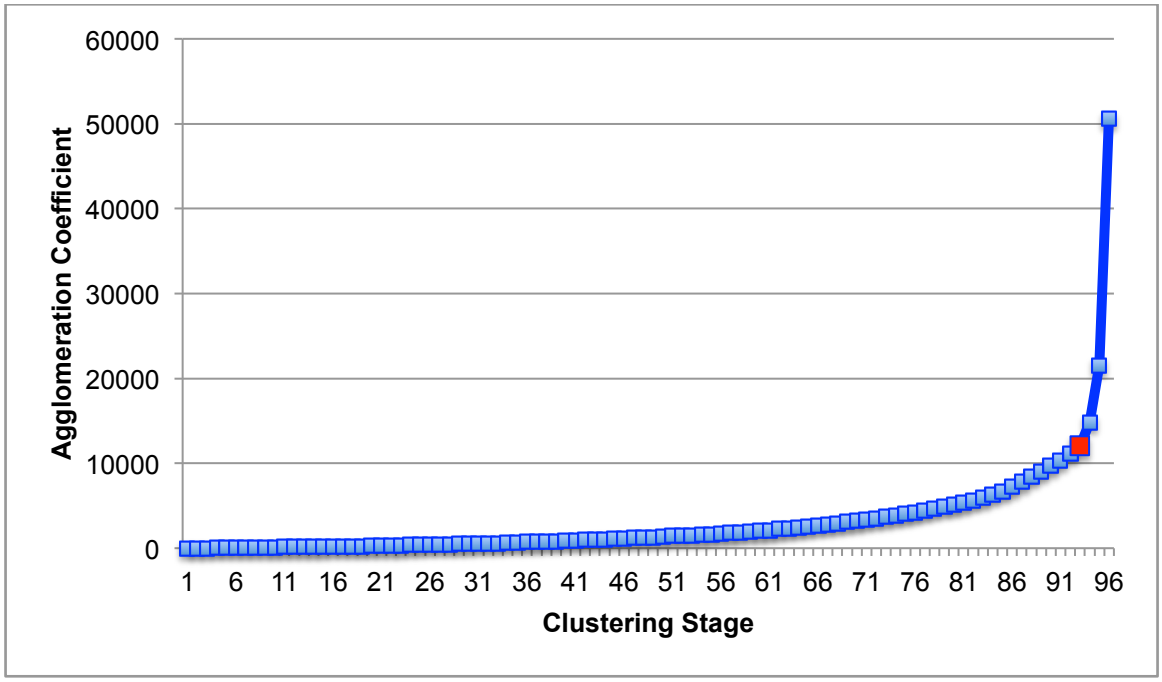


*Figure 15.* ***Scree Plot of Agglomeration Coefficients for the START Strength Scale Scores***

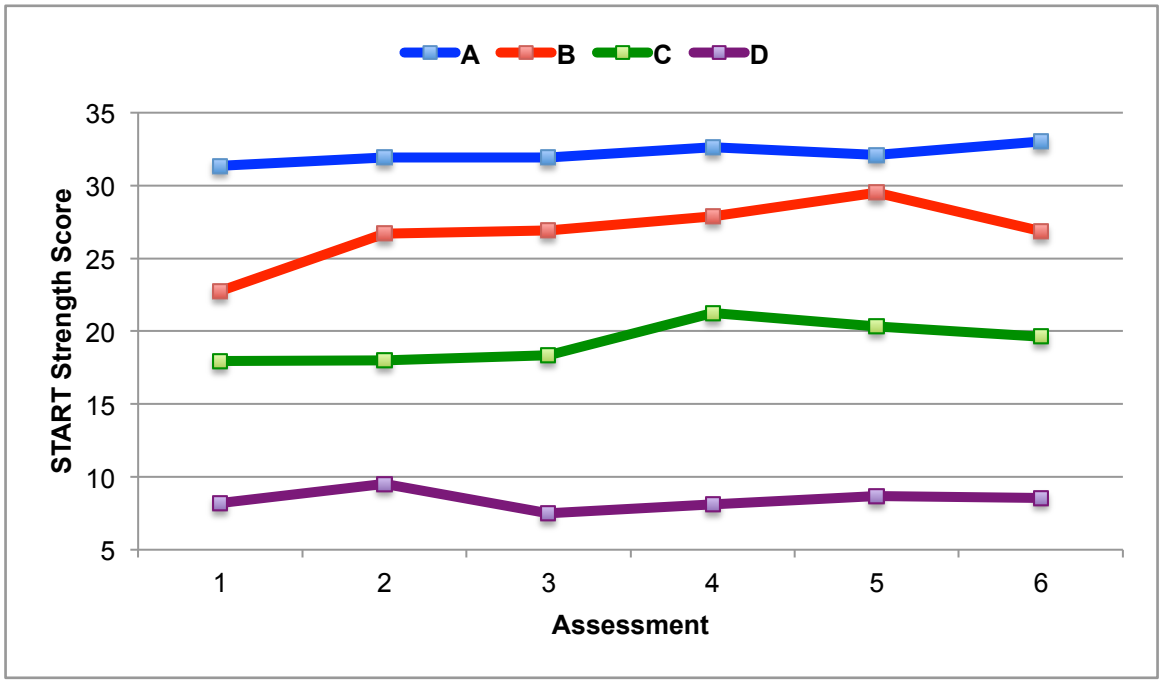


*Figure 16.* ***Mean START Strength Scale Scores across Six Assessment Separated by Cluster***

Overall, four clusters were identified on each of the dynamic scales. The four clusters appear to generally correspond with each other across the scales based mainly on the relative starting point of the participant's scores and pattern of change. That is, with some variability and reversing the pattern for the START Strength scale, the following four clusters were identified: one group started with low scores and gradually decreased or remained constant (Cluster A), one group started with scores in the middle range and gradually decreased or remained constant (Cluster B), one group that started with higher scores and then changed considerably (Cluster C), and a final group that started with high scores and remained high (Cluster D).

Subsequently, chi-square analyses were performed to determine the extent of overlap between the clusters identified for each of the dynamic scales. Examining the HCR-20 C scale clusters and the HCR-20 R scale clusters, an association was found in cluster memberships, $\chi^2(9, N = 99) = 66.67$, $p < .001$, $C = .63$. The HCR-20 C scale clusters were also associated with both the START Vulnerability clusters, $\chi^2(9, N = 97) = 71.67$, $p < .001$, $C = .65$, and the START Strength clusters, $\chi^2(9, N = 97) = 36.20$, $p < .001$, $C = .52$. Furthermore, the HCR-20 R scale clusters were associated with both the START Vulnerability clusters, $\chi^2(9, N = 97) = 83.92$, $p < .001$, $C = .68$, and the START Strength clusters, $\chi^2(9, N = 97) = 45.85$, $p < .001$, $C = .57$. Finally, an association was found between cluster membership on the START Vulnerability scale and the START Strength scale, $\chi^2(9, N = 97) = 62.90$, $p < .001$, $C = .63$.

Table 6 presents the percent of overlap in participant's membership in corresponding clusters between the four dynamic scales. As seen in Table 6, a substantial proportion of individuals were in corresponding clusters, in decreasing frequency, on the HCR-20 R scale and the START Vulnerability scale (54.6%), the START Vulnerability and START Strength scales (52.6%), the HCR-20 C scale and the START Vulnerability scale (51.5%), the HCR-20 R scale and the START Strength scale (46.4%), the HCR-20 C and HCR-20 R scales (44.4%), and finally the HCR-20 C scale and the START Strength scale (37.1%). Table 6 presents the percent overlap for each of the clusters using each of the different scale's clusters as the reference, as the number of participants in the different clusters varied across the scales. The largest overlap was seen in the cluster of individuals that remained low throughout the

assessments (Cluster A) and the cluster that remained high throughout the assessments (Cluster D), whereas less overlap was seen in the remaining clusters.

***Table 6. Overlap in Cluster Membership across Dynamic Scales***

| | Percent Agreement | | |
|---|---|---|---|
| **HCR-20 Clinical Clusters** | **HCR-20 Risk Management** | **START Vulnerability** | **START Strength** |
| **A** | 61.8% | 78.8% | 45.5% |
| **B** | 37.5% | 58.1% | 32.3% |
| **C** | 12.5% | 0.0% | 25.0% |
| **D** | 52.9% | 35.3% | 41.2% |
| **Total** | 44.4% | 51.5% | 37.1% |
| **HCR-20 Risk Management Clusters** | **HCR-20 Clinical** | **START Vulnerability** | **START Strength** |
| **A** | 70.0% | 82.8% | 58.6% |
| **B** | 36.4% | 57.6% | 45.5% |
| **C** | 8.0% | 16.6% | 29.2% |
| **D** | 81.8% | 54.5% | 54.5% |
| **Total** | 44.4% | 54.6% | 46.4% |
| **START Vulnerability Clusters** | **HCR-20 Clinical** | **HCR-20 Risk Management** | **START Strength** |
| **A** | 60.5% | 55.8% | 55.8% |
| **B** | 42.9% | 45.2% | 42.9% |
| **C** | 0.0% | 66.7% | 66.7% |
| **D** | 100.0% | 100% | 83.3% |
| **Total** | 51.5% | 54.6% | 52.6% |
| **START Strength Clusters** | **HCR-20 Clinical** | **HCR-20 Risk Management** | **START Vulnerability** |
| **A** | 57.7% | 65.4% | 92.3% |
| **B** | 31.3% | 46.9% | 56.3% |
| **C** | 17.4% | 30.4% | 17.4% |
| **D** | 43.8% | 37.5% | 31.3% |
| **Total** | 37.1% | 46.4% | 52.6% |

*Note.* Values provided are percent overlap or agreement in corresponding clusters.

Table 7 presents results examining the correspondence between participants membership in clusters across all four dynamic scales, as opposed to the dichotomous comparisons presented in Table 6. As seen in table 7, 19.6% of participants were consistently in corresponding clusters across all four of the dynamic scales. The majority of the sample (51.5%) was grouped in two different clusters across the four scales. Examining the combination of cluster membership, a notable proportion of the sample (39.2%) was grouped in the same cluster on three of the dynamic scales, only diverging in their cluster membership on the fourth scale, and a smaller proportion (13.4%) was grouped in the same cluster on two of the scales and a different cluster on the other two scales. Only 1% of participants were grouped in all four clusters, one for each scale. Table 7 also presents the farthest distance between clusters in which each participant was included, in the sense that Cluster A is one unit away from Cluster B, two units away from Cluster C, and three units away from Cluster D, and so on. Once again, the majority of participants were either consistently in the same cluster on all four scale (19.6%), or in adjacent clusters on all the scales (49.5%). Membership in non-adjacent clusters (i.e., A and C, or B and D) was less common and only 3.1% were grouped in the most disparate clusters on different scales (i.e., A and D).

***Table 7. Frequencies of Cluster Memberships across all Dynamic Scales***

| | **Number of Different Clusters of Participant Membership** | | | |
|---|---|---|---|---|
| | One | Two | Three | Four |
| **Percent** | 19.6% | 51.5% | 27.8% | 1.0% |

| | **Combination of Clusters of Participant Membership** | | | | |
|---|---|---|---|---|---|
| | All Four the Same | Three the Same | Two the Same Two the Same | Two the Same Two Different | All Four Different |
| **Percent** | 19.6% | 39.2% | 13.4% | 26.8% | 1.0% |

| | **Farthest Cluster Disparity in Participant Membership** | | | |
|---|---|---|---|---|
| | Zero | One | Two | Three |
| **Percent** | 19.6% | 49.5% | 27.8% | 3.1% |

*Note.* $N = 97$. Values provided are percent of participants.

## Research Question 5: Association of Change with Future Violence

Finally, in order to address the fifth research question regarding the association of change on dynamic scales with subsequent violence, GEE was used. In each of the following analyses, change in the dynamic scale scores from one assessment to the next were entered separately into bivariate GEE analysis. The results of these analyses are displayed in Table 8. Pertaining to the perpetration of any violence, change in the HCR-20 C scale scores and change in the HCR-20 R scale scores over multiple assessments were predictive of subsequent violence. An odds ratio of 1.25 for the C scale indicates that on average a one-unit increase in the C scale score from one assessment to the next is associated with a 25% increase in the likelihood of violence. In comparison, a one-unit increase in the R scale from one assessment to the next was associated with a 21% increase in the likelihood of violence. Changes in either of the START scales across assessments were not predictive of the broad definition of violence.

***Table 8. Association of Change in Dynamic Scales with Future Violence***

| | Broad Violence | | | Narrow Violence | | |
|---|---|---|---|---|---|---|
| | **OR** | **95% CI** | ***p*** | **OR** | **95% CI** | ***p*** |
| **HCR-20** | | | | | | |
| C Scale | 1.25 | [1.10, 1.41] | .001 | 1.33 | [1.14, 1.55] | <.001 |
| R Scale | 1.21 | [1.06, 1.38] | .004 | 1.23 | [1.05, 1.44] | .009 |
| **START** | | | | | | |
| Vulnerability | 1.03 | [0.98, 1.08] | .311 | 1.05 | [0.98, 1.12] | .198 |
| Strength | 0.97 | [0.92, 1.01] | .148 | 0.94 | [0.90, 0.97] | .001 |

Pertaining to the perpetration of more serious violence (i.e., only those violent acts included in the MacArthur Community Violence Interview), once again change in both the HCR-20 C and R scale scores over multiple assessments were associated with future violence. Moreover, change in START Strength scores was also predictive of future serious violence. In this case, an odds ratio of 0.94 indicates that on average a one-unit increase in START Strength scores from one assessment to the next is

associated with a 6% decrease in the odds of violence. However, change in the START Vulnerability scale was not associated with subsequent violence.

As seen in Table 8, changes in the HCR-20 C scale showed the strongest and most consistent relationship with future violence by either definition, followed by the HCR-20 R scale scores, and then the START Strength scores. As well, changes in the dynamic scales were slightly more strongly associated with more serious forms of violence. That is, the odds ratios are consistently, but only slightly, farther from zero when considering the narrow definition of violence compared to the broad definition of violence. For instance, a one-unit increase in the HCR-20 C scale at reassessment is associated with a 25% increase in the likelihood of any violence and a 33% increase in the likelihood of more serious forms of violence.

In addition, these analyses were conducted with sample type (civil psychiatric versus correctional) as a covariate. Controlling for sample type did not change the overall pattern of results, with changes in the HCR-20 C and R scales associated with both types of violence and change in the START Strength scale associated with serious violence. Adding this covariate changed the reported odds ratios minimally (from no change to a maximum change of 0.02), and in each case any change increased the strength of the association. Thus, it appears as though the present findings are robust across the two subsamples.

For comparison purposes, Table 9 includes the results of GEE analyses using the scale scores at each of the assessment periods as the predictors, as opposed to the change scores from one assessment to the next. These results reveal that each of the dynamic scales is predictive of subsequent violence by either definition. Moreover, the scales appear to be equally predictive of both types of violence (i.e., relatively no change in odds ratios between the broad and narrow definitions of violence). These analyses were also conducted controlling for sample type. Once again, no change in the overall pattern of association was found and the odds ratios changed minimally (from no change to a maximum change of 0.04).

***Table 9. Association of Dynamic Scales with Future Violence***

| | Broad Violence | | | Narrow Violence | | |
|---|---|---|---|---|---|---|
| | **OR** | **95% CI** | ***p*** | **OR** | **95% CI** | ***p*** |
| **HCR-20** | | | | | | |
| C Scale | 1.47 | [1.34, 1.62] | <.001 | 1.47 | [1.33, 1.63] | <.001 |
| R Scale | 1.45 | [1.33, 1.59] | <.001 | 1.41 | [1.27, 1.56] | <.001 |
| **START** | | | | | | |
| Vulnerability | 1.12 | [1.09, 1.15] | <.001 | 1.12 | [1.08, 1.16] | <.001 |
| Strength | 0.92 | [0.90, 0.95] | <.001 | 0.91 | [0.89, 0.94] | <.001 |

# Discussion

## Limitations and Clarifications

Prior to discussing the main findings and implications of these findings, it is imperative to make clear the limitations of the present study. First, the combination of data from two subsamples was not ideal. Had sufficient data been available for each of the subsamples across the six assessments then the analyses would have been conducted separately within each of the two subsamples. Due to the exploratory nature of the research questions and analytic approaches, as well as the overall goal of determining whether it was possible for the dynamic scales to change over time, the samples were combined to provide a greater number of observations. This was done in order to provide a larger sample for the exploratory analyses, as well as to increase power for the statistical procedures. Meaningful differences exist between psychiatric inpatients and correctional offenders that may impact the amount, rate, and type of change seen on these sorts of dynamic scales. On the other hand, meaningful differences also exist amongst both psychiatric inpatients and correctional offenders that also likely impact the particular intraindividual trajectories of change on these scales. Additionally, sample type was controlled for in certain analyses, which revealed that the findings were robust across the two subsamples. Nevertheless, this limitation should be kept in mind when considering the generalizability of the current findings to unique samples or populations of interest, yet it does not prevent conclusions from being drawn regarding the general characteristics of change in dynamic risk scales over time in the present study.

In addition, a second limitation of this study was the high attrition rate and the resulting missing data that reduced the number of observations included in many of the analyses. Attrition was seen at each of the assessment intervals; however, roughly three-quarters of the sample had at least one follow-up and nearly half of the sample was assessed the full six times. The rate of attrition in the present study was

comparable to the MacArthur study of mental disorder and violence (Monahan et al., 2001). This study, often considered the gold standard for longitudinal risk assessment studies, also included six assessments and a total of 49.6% of participants completed all of the assessments, compared to 46.3% in the present study. Nevertheless, the number of observations included in the RCI analyses would have increased substantially with full data from every participant (for 622 observations to 1175). As such, the reported proportions of reassessments that changed on each of the dynamic scales may have been quite different if all participants were assessed at each assessment timeframe. Additionally, the statistical results reported for the other analyses also include a degree of error associated with the lack of complete data for many participants. As discussed in more detail above, missing data in the present study was mainly handled through listwise deletion. Multiple imputation was not considered appropriate due to the exploratory nature of the analyses and the chosen analytic strategies. The combination of multiple imputation with the selected analytic approaches would have likely led to similar error rates and may have masked some degree of change seen in those with complete or partially complete data.

Moreover, although no differences were found on important demographic variables or baseline scores on the dynamic risk scales between those who did not complete any follow-ups and those who completed at least one follow-up, some differences were found between those that completed all six assessments and those who did not. Specifically, compared to participants that did not complete all six assessments, participants who completed all six assessments were found to differ in their baseline scores on the HCR-20 C scale, START Vulnerability scale, and START Strength scale. These findings have the most direct implications on the analyses examining the overall type of change and the identification of subgroups that change according to different trajectories. Only participants with data from all six assessments were included in these analyses; thus, these results should be interpreted with caution. Nevertheless, despite the rate of attrition in the present study, the results from the available data can still shed some light on the research questions of interest.

Another limitation concerns the level of analysis. The current study focused exclusively on the scale level, as opposed to the individual item or risk factor level. Focusing on the scale level may have masked some change occurring at the item level.

It may be the case that individual items (risk or strength factors) changed more frequently than was observed based on the scale level analyses. Specific items may have changed in opposite directions, thus, cancelling each other's effects when examined only at the scale level. However, it is much more likely that items would change in similar directions. The items included on these scales are grouped heuristically, thematically, and temporally. Many of the risk factors on these scales can influence the other risk factors, thus, making it unlikely to see change in opposite directions. For instance, with regards to the HCR-20, an individual who becomes noncompliant with treatment, thus, increasing their score on unresponsive to treatment, is more likely to experience a psychotic break, increasing their score on active symptoms of major mental illness, which will likely reduce their level of insight, increasing their score on lack of insight, and increase their impulsiveness, increasing their score on impulsivity.

In addition to this theoretical rationale, the items included in these scales are also statistically associated. For instance, in the present study item inter-correlations were calculated based on the baseline scores on each of the dynamic scales. With regards to the HCR-20 C scale, of a possible 10 unique item associations, seven positive inter-correlations were found (*MIC* = .20, range = -.21 to .45). On the HCR-20 R scale, of a possible 10 unique associations, eight positive inter-correlations were present (*MIC* = .26, range = -.14 to .44). With regards to the START Vulnerability scale, of a possible 190 unique item associations, 121 positive inter-correlations were found (*MIC* = .21, range = .14 to .65). Finally, on the START Strength scale, 174 unique item inter-correlations were found out of the possible 190 (*MIC* = .31, range = -.01 to .62). Nevertheless, it is possible, but unlikely, to see change in opposite directions occurring on certain items on the HCR-20 and START. As such, any conclusions drawn from the current study must be qualified when considered the particular items on these scales.

Another potential limitation, and point to clarify, concerns the chosen analytic strategies. The primary concern in this area is change within a given individual over time (i.e., intraindividual change). Many of the analytic and statistical procedures often employed in these lines of inquiry use aggregate data and results to draw conclusions about particular individuals. For instance, several authors have used aggregate level, mean change on a putatively dynamic scale as evidence of intraindividual change (e.g.,

Belfrage & Douglas, 2002; Holliday et al., 2012; Neves et al., 2010). It is important to make clear when the specific question of interest and analytic approach concerns the intraindividual level or interindividual level. In the present study, a combination of approaches were undertaken to address the five research questions. Research questions one and two, regarding whether change is seen on these scales and whether the reassessment length affects this change, were investigated using change scores and RCI analyses. These analyses speak to intraindividual change. In comparison, the final three research questions, regarding the type of change, groups that change differently, and the association of change with violence, used aggregate (group) level data and analytic approaches. With these limitations and clarifications in mind, several conclusions can be drawn from the current results.

## Are Dynamic Risk Scales Truly Dynamic?

The purpose of the present study was to investigate change and characteristics of change in (putatively) dynamic scales on structured risk assessments instruments over time. The primary question of interest concerned whether putatively dynamic scales on the HCR-20 and START do in fact change over time. Overall, the present study found that intraindividual change is seen in a considerable proportion of reassessments. Change was seen in the majority of reassessments on all four of the dynamic scales of interest, with a slightly greater proportion of intraindividual change seen on the START scales than the dynamic scales on the HCR-20. However, only a small proportion of the reassessments reached the level of reliable change according to the RCI analyses (from 5.2% to 6.6% depending on the dynamic scale). This may be due to the fact that high variability on the dynamic scales in this sample made it difficult to reach the threshold level of reliable change. Moreover, even small amounts of change on one of these scales can be very meaningful clinically. For instance, a decrease of two-points on the HCR-20 C scale may be very meaningful in practice if this reduction meant that the individual is no longer experiencing acute psychotic symptoms. Although only a small proportion of the reassessment met the threshold level of reliable change, intraindividual change was seen in a much greater proportion of the assessments. Accordingly, it may be concluded that at least some of the items (risk and

strength factors) included on the dynamic scales of the HCR-20 and START are in fact dynamic risk (or strength) factors according to the typologies described above.

The results of the current study coincide with the available research. For instance, investigating the START:AV, Viljoen and colleagues (2012) found reliable intraindividual change on 17.5% of the START Strength scores and 15.6% of the START Vulnerability scores. Comparatively, the current study found reliable intraindividual change in 6.6% and 6.3% of START Strength and Vulnerability scores, respectively. As well, Draycott and colleagues (2012) examined change on the HCR-20 using RCI analyses. These authors found intraindividual change on 31.0% and 17.2% of the reassessments on the HCR-20 C and R scales, respectively. Comparatively, the current study found reliable change in 5.2% and 6.6% of the HCR-20 C and R scales, respectively. Thus, less reliable change was seen in the current study; however, this may be explained by the difference in reassessment length (discussed in more detail below) or sample characteristics. A greater proportion of reliable change may have been seen by Viljoen and associates (2012), as they used a sample of adolescents and more change may be expected in adolescents compared to adults. As well, Draycott et al. (2012) had a relatively small sample size of 29 participants and this may have impacted the amount of reliable change. Finally, Michel and colleagues (2013) reported the raw percentages of participants whose scores changed between any five assessments. They found that 95.4% of participants' scores changed on the HCR-20 C scale and 91.5% of participants' scores changed on the HCR-20 R scale. In the present study, change was seen in 74% and 73.3% of reassessments on the HCR-20 C and R scales, respectively. However, Michel and colleagues reported the percentage of participants whose scores changed between at least one of the five assessments, compared to the current percentages that report the overall number of reassessments that changed from each assessment to the next.

Next, the rate of change was investigated by examining whether the proportions of assessments that display change varied depending on the length of the reassessment interval. Overall, for the HCR-20, the current study found that less change was seen when the reassessment interval was shorter compared to a longer reassessment interval. With regards to the HCR-20 C scale, when the reassessment interval was 30 days or less, 72% of the scores changed with 2.8% reaching the level of reliable change.

When the reassessment interval was 60 days or more, 80% of the scores changed with 7.4% reaching the level of reliable change. This pattern may explain the larger amounts of reliable change seen in the previous empirical research, such as Draycott and colleagues (2012) who found reliable change on 31.0% and 17.2% of the reassessments on the HCR-20 C and R scales, respectively, over an 18-month reassessment interval. The current study may, thus, have found higher rates of reliable change if the reassessment interval had been longer.

In contrast, the proportion of change on the START scales in the current study did not appear to vary as a function of the reassessment interval. However, once again it may be the case that if the reassessment interval had been extended, more change may have been seen on the START scales. For example, over a three-month period, Viljoen et al. (2012) saw reliable change on 17.5% of the START Strength scores and 15.6% of the START Vulnerability scores, compared to 6.6% and 6.3% in the current study. Thus, it may be the case that the different length intervals investigated in the present study were too constricted to detect differential change on the START scales over different assessment lengths. That is, different proportions of intraindividual change may have been seen if comparisons were made between reassessments of 30 days or less compared to six months or more, or some other much longer reassessment interval.

## How do Dynamic Scales Change over Time?

Subsequently, the type (i.e., shape or pattern) of change was investigated. Overall, aggregate (mean) changes were seen on all four of the dynamic scales across the six assessments. Mean decreases were found on the three risk scales, while a mean increase was seen on the START Strength scale. Similar group level mean changes were found on the HCR-20 C and R scales by Belfrage and Douglas (2002) and Michel et al. (2013); as well, Douglas and colleagues (2011) found mean decreases on the HCR-20 C scale across assessments. Other authors have found mean decreases on additional risk assessments instruments. For instance, Schlager and Pacheco (2011) found a mean decrease in total LSI-R scores, as well as seven of the 10 subcomponents. Similarly, Holliday and colleagues (2012) found a mean decrease in

total LS/CMI scores, as well as mean decreases on all four criminogenic need items. However, one study by Neves et al. (2010) found no group level change on the HCR-20 dynamic scales.

Douglas and colleagues (2011) are the only authors to report the type or shape of this group level change. These authors found a linear decrease in HCR-20 C scale scores across five assessments. The current study found that both a linear and a quadratic trend were fitting of the mean pattern of change over time on the HCR-20 C scale, while a linear trend was fitting of the mean pattern of change on the HCR-20 R scale. The shape of change on the START scales was much less clear with linear, quadratic, and cubic trends fitting the START Strength scale and linear and quadratic trends fitting the pattern seen on the START Vulnerability scale. These findings highlight the fact that group level change on these instruments is complex, and at the intraindividual level, individuals likely change according to different trajectories.

As such, the next research question investigated subgroups of individuals that change differently on the risk assessment scales. On each of the four dynamic scales, four different clusters, or groups of individuals, were found based on their pattern of change across the assessments. Across all four of the dynamic scales (reversing the pattern seen on the START Strength scale to coincide with the risk scales), there was generally one group that started with low scores and decreased slightly over time, and another group that started with high scores and remained fairly constant over time. There was also generally a group that started in the middle range then changed during the first few assessments before levelling off for the remainder of the assessments. The specific trajectories of the final group varied across each of the dynamic scales.

Only one study has examined whether individuals change differently on dynamic risk scales. Douglas et al. (2011) found five groups based on their pattern of change across four assessments on the HCR-20 C scale. These authors also found one group of individuals that started with low scores and remained low over time. However, the other groups they identified appear to follow different patterns than those seen in the present study. This fact is not surprising, as the manner in which individuals, and groups of individuals, change is most likely dependent on a number of characteristics, such as the type and severity of mental illness or the intensity of intervention strategies. Thus,

the patterns identified in the current study are not necessarily thought to be generalizable to additional samples. Nevertheless, it is still important to be aware that there are groups of individuals that change in different manners on the different scales and change is likely affected by a number of individual characteristics that result in particular trajectories across assessments.

## Are Dynamic Risk Factors actually Causal Risk Factors?

Finally, and possibly most importantly, the association between change on the dynamic scales and violence was investigated. The present study found that regardless of the definition of violence, change on the HCR-20 dynamic scales was associated with violence. In comparison, Michel and colleagues (2013) found that change on the HCR-20 R scale was associated with future violence, as were changes on three of the C items and three of the R items. As well, Wilson and colleagues (2013) found that change on the HCR-20 C and R scales were both predictive of future violence, even after controlling for the HCR-20 static items (i.e., the Historical scale). Thus, for the HCR-20 the evidence is beginning to accumulate that the dynamic scales do in fact change over time and this change is associated with violence. From this evidence, it appears that at least some of the items on the HCR-20 C and R scales may best be described as, not only dynamic risk factors, but causal risk factors. However, far more evidence is needed to make this claim with conviction.

With regards to the START scales, the present study found that change on the START Strength scale was associated with serious violence only, but not the broad definition of violence. Change on the START Vulnerability scale was not associated with either operationalization of violence. In comparison, Wilson and colleagues (2013) found that change on the START Vulnerability scale, but not the START Strength scale, was associated with violence. This association was also still present after controlling from the HCR-20 static items. The available evidence for the START appears to be less clear that that for the HCR-20. The two studies that have investigated this issue each found contradictory results in that change on only one of the START scales was associated with violence. As such, more empirical evidence is needed to draw conclusions regarding the association of change on the START scales with violence.

# Implications for Risk Management

## *Ongoing Monitoring and Assessment*

Current practice is much more focused on ongoing monitoring and assessment than in the past (Doyle & Logan, 2012; Mulvey & Lidz, 1995; Douglas & Skeem, 2005; Skeem et al., 2000; Webster et al., 2000). Short-term prediction and implementation of corresponding risk management strategies are increasingly being stressed (Andrews & Dowden, 2007; Doyle & Logan, 2012; Hanson & Harris, 2000). As such, there is an ever-increasing need to understand the dynamic nature of risk. Currently, structured assessment instruments provide theoretical guidelines and rational for aspects of dynamic risk, including the appropriate time interval to gather information from for rating dynamic risk factors, the appropriate reassessment interval, and the appropriate time to intervene. For instance, with regards to the HCR-20, Douglas and Reeves (2010) provide general guidelines about the appropriate length of time to consider when making ratings on dynamic items and the appropriate reassessment interval. In general, high-risk individuals should be reassessed more frequently and assessment should be conducted whenever any significant changes have occurred in an individual's life.

The available empirical research has used various reassessment intervals from months to years; therefore, at present few conclusions can be drawn from the empirical data regarding the frequency or rate of change seen in dynamic scales and items. The present study found that the START scales appear to capture a similar amount of intraindividual change regardless of the reassessment interval; however, other authors have found much higher proportions of intraindividual change using longer reassessment intervals (Viljoen et al., 2012). In contrast, for the HCR-20, the present study found that the intraindividual change is seen in greater proportions of individuals as the length of the reassessment interval increases. However, at this time little is known about the rate of change over time for various risk factors, or the ability to classify certain risk factors as acute versus stable dynamic risk factors. Empirically, change has been seen to occur in some factors over weeks, while change has not been seen in some cases over 18 months (e.g., Holliday et al., 2012; Draycott et al., 2012). Thus, more longitudinal empirical work is needed in this area to better understand the ability of

dynamic risk factors to change over time and the factors that affect the rate of change within and across individuals.

Additionally, empirical findings regarding the proportions of individuals that change in a given period of time are only one factor to consider when determining the appropriate monitoring strategy. Risk state "changes over time, place and circumstance" (Craig et al., 2005, p. 79). Thus, a better understanding of the fluctuations in risk state and levels of various dynamic factors will allow practitioners to make more informed decisions about changes to supervision and management plans, whether an individual is responding to intervention strategies, and when to introduce interventions (Douglas & Skeem, 2005). Practitioners must not only have a sense of the rapidity and frequency of change in dynamic risk factors, but also the individual characteristics that are capable of affecting the rate of change on dynamic factors in order to establish a defensible schedule for assessing and monitoring a given individual (Douglas & Skeem, 2005). Accordingly, numerous issues must be taken into consideration when determining how often to assess a particular individual with a particular tool; amongst other considerations, both characteristics of the individual and the assessment instrument will impact the appropriate reassessment length.

### *Targeted Intervention Strategies*

Best practice in risk assessment and effective risk management necessitate consideration of dynamic risk factors (Andrews & Bonta, 2010a, 2010b; Craig et al., 2005; Wilson et al., 2013). Assessment of dynamic risk factors is meant to identify targets of treatment and case formulizations can serve as blueprints for the provision of targeted intervention strategies that are most likely to reduce violence (Bonta, 2002). Numerous studies have shown that interventions targeting supposed criminogenic needs are more successful at reducing recidivism (e.g., Andrews et al., 1990; Andrews & Bonta, 2010a, 2010b). Accordingly, the accurate identification of dynamic risk factors in each individual could allow for targeted interventions that are idiographically chosen to be maximally effective. Moreover, repeated assessments of dynamic risk factors may allow practitioners to monitor the effectiveness of these targeted interventions (Bonta, 2002).

The current study found that change on the dynamic scales of the HCR-20 is associated with violence. As such, it may be appropriate to label some of the risk factors included on these scales as causal dynamic risk factors according to the Kraemer and colleagues' (1997) typology. The identification of this type of risk factor is paramount in order to succeed in the ultimate goal of most violence risk assessments (i.e., the prevention of future violence). Additional empirical research has indicated that structured assessment instruments are able to monitor treatment progress, and these instruments may in fact contain items that are able to measure and monitor causal risk factors (Douglas et al., 2011; Draycott et al., 2012; Holliday et al., 2012; Michel et al., 2013; Schlager & Pacheco, 2011; Wilson et al., 2013). More research is needed in this area to determine which specific risk factors on structured risk assessment instruments are best characterized as causal risk factors (or criminogenic needs when referring to risk of general criminality) in order to target interventions appropriately.

## Combining SPJ and RNR Approaches

More generally, the above discussions call attention to the utility of the SPJ and RNR models. These models share a number of features and can be considered complimentary to one another (Guy, 2008). Tools developed under these two models have been shown to evidence comparable rates of predictive validity (Campbell et al., 2009). This fact is not surprising given there is a great deal of content overlap (i.e. risk factors) between instruments developed under each model (e.g., the HCR-20 and LSI tools). As well, both models stress the importance and necessity of assessing dynamic risk factors in addition to static risk factors (Andrews & Bonta, 2010a, 2010b; Douglas & Kropp, 2002; Guy, 2008). Under both models static risk factors are included as they speak to the individual's overall risk status and, thus, can distinguish groups of individuals that require more intense risk management interventions. On the other hand, both models necessitate the assessment of dynamic risk factors to determine the individual's risk state and appropriate targets for intervention. Finally, and possibly most fundamentally, both of these models stress the profound relationship between risk assessment and risk management (Andrews & Bonta, 2010a, 2010b; Douglas et al., 2013; Douglas & Kropp, 2002; Guy, 2008). These models both consider a comprehensive risk assessment as a means to the larger and ultimate goal of

implementing risk management strategies to mitigate risk and prevent offending or violence. Due to the fundamental similarities between these models, their combined use may be appropriate. Using the SPJ approach to risk assessment and the RNR approach to risk management may result in an optimal approach to reducing crime and violence.

## Conclusions and Future Directions

Scholars have called attention to the lack of empirical evidence regarding (a) the changeability of hypothetically dynamic risk factors, (b) the functional relationships of risk factors with violence and the roles that risk factors play in the perpetration of violence, (c) the changeability of global assessment of risk state, and (d) the relation between observed fluctuations in dynamic risk factors and corresponding fluctuations in the propensity for violence (Douglas & Skeem, 2005). The present study adds to the limited body of empirical research that can speak to these issues. Specifically, the present study adds to the emerging empirical evidence that the dynamic scales on the HRC-20 and START are in fact dynamic. That is, these scales are capable of capturing intraindividual change over time. In addition, the present study found that the length of the reassessment interval impacts the proportion of intraindividual change seen on the HCR-20 but not the START. The specific trajectory (or pattern) of change seen on each of the dynamic scales was found to vary across individuals with four groups emerging on each scale that change according to different patterns. Finally, change on the HCR-20 was associated with both forms of violence and change on the START Strength scale was associated with serious violence only.

As it is believed that the demand for risk assessments will proliferate, the underlying science must continue at an increased speed (Douglas & Kropp, 2002). A major task and challenge facing the field is the development and evaluation of empirically sound approaches for assessing the variable aspects of violence risk (Douglas & Skeem, 2005; Dvoskin & Heilbrun, 2001). The majority of work in this area has been theoretical in nature and there is currently limited empirical knowledge regarding the rapidity and frequency of change in risk factors over time (Douglas & Skeem, 2005). Future research is needed to accurately categorize risk factors

according to the various available typologies, in terms of their rates of change over time and their types of association with violence. As well, research is needed to determine the temporal precedence of risk factors in relation to other risk factors and to violence, the associations between risk factors, and the predictive strength of dynamic risk factors alone and in combination, as well as "to disentangle the independent, indirect, interactive, or transactional effects of dynamic risk factors on violence" (Douglas & Skeem, 2005, p. 368).

Essential elements of future research include a prospective repeated measures design that assesses hypothetically dynamic risk factors at enough time points to examine patterns and frequency of change, as well as frequent measurement of violence in order to determine the effect of change in risk factor levels on violence (Brown et al., 2009; Douglas & Kropp, 2002; Douglas & Skeem, 2005; Jones et al., 2010). Future research should allow for testing any mediating or moderating effects on the relationship between dynamic risk factors and violence (Douglas & Skeem, 2005). Ideally, future studies will focus on issues of generalizability related to different settings and populations, as change is likely quite different across individuals and groups. The influence of important individual characteristics must also be examined in relation to the changeability of dynamic risk factors and their relation to violence. Finally, future research must use appropriate statistical techniques that can assess the complex nature of change at the idiographic level.

# References


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

Andrews, D. A., & Bonta, J. (2010a). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*(1), 39-55.

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

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*(1), 19-52.

Andrews, D. A., Bonta, J., & Wormith, S. (2004). *The Level of Service/Case Management Inventory user's manual*. Toronto, Canada: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, S. (2010). The Level of Service (LS) assessment of adults and older adolescents. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 199-225). New York, NY: Routledge/Taylor & Francis Group.

Andrews, D. A., & Dowden, C. (2007). The Risk-Need-Responsivity model of assessment and human service in prevention and corrections: Crime-prevention jurisprudence. *Canadian Journal of Criminology and Criminal Justice, 49*, 439-464.

Bartko, J. J., Carpenter, W. T. (1976). On the methods and theory of reliability. *Journal of Nervous and Mental Disease*, 163, 307-317.

Beech, A., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A Journal of research and Treatment, 14*(2), 155-167.

Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender version. *Sexual Abuse: A Journal of research and Treatment, 22*(2), 234-251.

Beggs, S. M., & Grace R. 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.

Belfrage, H., & Douglas, K. S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health, 1*(1), 25-36.

Belfrage, H., Fransson, G., & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *Journal of Forensic Psychiatry, 11*(1), 167-175.

Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior, 29*(4), 355-379.

Braithwaite, E., Charette, Y., Crocker, A. G., Reyes, A. (2010). The predictive validity of clinical ratings of the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health, 9*(4), 271-281.

Brown, S. L., Amand, M. D., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: A three-way prospective study. *Law and Human Behaviour, 33*, 25-45.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36*(6), 567-590.

Chu, C. M., Thomas, S. D. M., Ogloff, J. R. P., & Daffern, M. (2011). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. *International Journal of Forensic Mental Health, 10*(4), 337-345.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137.

Cooke, D. J., & Michie, C. (2013). Violence risk assessment: From prediction to understanding – or from what? To why? In C. Logan & L. Johnstone (Eds.), *Managing clinical risk: A guide to effective practice* (pp. 3-25). New York, NY: Routledge/Taylor & Francis Group.

Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression, 11*(1), 65-84.

Craissati, J., & Beech, A. (2003). A review of dynamic variables and their relationship to risk prediction in sex offenders. *Journal of Sexual Aggression, 9*(1), 41-55.

Daffern, M., & Howells, K. (2007). The prediction of imminent aggression and self-harm in personality disordered patients in a high security hospital using the HCR-20 Clinical scale and the Dynamic Appraisal of Situational Aggression. *International Journal of Forensic Mental health, 6*(2), 137-143.

Daffern, M., Howells, K., Hamilton, L., Mannion, A., Howard, R., & Lilly, M. (2009). The impact of structured risk assessments followed by management recommendations on aggression in patients with personality disorder. *The Journal of Forensic Psychiatry & Psychology, 20*(5), 661-679.

Dempster, R. J., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and Treatment, 14*(2), 121-138.

Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment, 24*(3), 685-700.

Diggle, P. J., Liang, K. Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford: Clarendon Press.

Dolan, M., & Fullam, R. (2007). The validity of the Violence Risk Scale second edition (VRS-2) in a British forensic inpatient sample. *The Journal of Forensic Psychiatry & Psychology, 18*(3), 381-393.

Douglas, K. S., Blanchard, A. J. E., Guy, L. S., Reeves, K. A., & Weir, J. (2010). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Available online: http://kdouglas.files.wordpress.com/2007/10/hcr-20-annotated-biblio-sept-2010.pdf

Douglas, K. S., Blanchard, A. J. E., & Hendry, M. C. (2013). Violence risk assessment and management: Putting structured professional judgment into practice. In C. Logan & L. Johnstone (Eds.), *Managing clinical risk: A guide to effective practice* (pp. 29-55). New York, NY: Routledge/Taylor & Francis Group.

Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology, 4*, 149-184.

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., & Eaves, D. (2008). *Historical, Clinical, Risk Management (Version 3): Assessing Risk for Violence.* Draft Version 1.0, November 14, 2008. Burnaby, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior, 29*(5), 617-658.

Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services, 54*(10), 1372-1379.

Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L. & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology, 67*(6), 917-930.

Douglas, K. S., & Reeves, K. (2010). The HCR-20 violence risk assessment scheme: Overview and review of the research. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 147-185). New York, NY: Routledge/Taylor & Francis Group.

Douglas, K., S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*(3), 347-383.

Douglas, K. S., Strand, S., & Belfrage, H. (2011, June). *Dynamic risk: Evaluating the nature and predictive validity of change on the Clinical and Risk Management scales of the HCR-20*. Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Barcelona, Spain.

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*(5), 479-510.

Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending. *Canadian Journal of Criminology, 42,* 449-467.

Doyle, M., & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services. *British Journal of Psychiatry, 189*, 520-526.

Doyle, M., & Logan, C. (2012). Operationalizing the assessment and management of violence risk in the short-term. *Behavioral Sciences and the Law, 30*, 406-419.

Draycott, S., Kirkpatrick, T., & Askari, R. (2012). An idiographic examination of patient progress in the treatment of dangerous and severe personality disorder: A reliable change index approach. *Journal of Forensic Psychiatry & Psychology, 23*(1), 108-124.

Dvoskin, J. A., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of the American Academy of Psychiatry and the Law, 29,* 6-10.

Dvoskin, J. A., & Steadman, H. J. (1994). Using intensive case management to reduce violence by mentally ill persons in the community. *Hospital and Community Psychiatry, 45*(7), 679-684.

Edens, J. F., Skeem, J. L., & Kennealy, P. J. (2009). The Psychopathy Checklist in the courtroom: Consensus and controversies. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 175-201). New York, NY: Guilford Press.

Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment, 24*(1), 5-28.

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabilitation Psychology, 56*(4), 267-288.

Garson, G. D. (2012). *Cluster analysis* (2012 Edition). Asheboro, NC: Statistical Associates Publishing, Blue Book Series.

Garson, G. D. (2012). *Missing values and data imputation* (2012 edition). Asheboro, NC: Statistical Associates Publishing, Blue Book Series.

Garson, G. D. (2013). *Generalized linear models / generalized estimating equations*. Asheboro, NC: Statistical Associates Publishing, Blue Book Series.

Grann, M., Sturidsson, K., Haggard-Grann, U., Hiscoke, U. L., Alm, P.-O., Dernevik, M., ... Woodhouse, A. (2005). Methodological development: Structured outcome assessment and community risk monitoring (SORM*). International Journal of Law and Psychiatry, 28*, 442-456.

Gray, N. S., Benson, R., Craig, R., Davies, H., Fitzgerald, S., Huckle, P., ... Snowden, R. J. (2011). The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *International Journal of Forensic Mental Health, 10*(4), 305-313.

Grittner, U., Gmel, G., Ripatti, S., Bloomfield, K., & Wicki, M. (2011). Missing value imputation in longitudinal measures of alcohol consumption. *International Journal of Methods in Psychiatric Research, 20*(1), 50-61.

Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey*. Unpublished dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.

Hanson, R. K. (1998). What do we know about sex offender risk assessment. *Psychology, Public Policy, and Law, 4*(1), 50-72.

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*(6)*,* 6-35.

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121-137.

Hart, S. D. (2001). Assessing and managing violence risk. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 13-25). Burnaby, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.

Hart, S. D. (2008). Preventing violence: The role o risk assessment and management. In A. C. Baldry, & F. W. Winkel (Eds.), *Intimate partner violence prevention and intervention: The risk assessment and management approach* (pp. 7-18). Hauppauge, NY: Nova Science.

Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior, 21*(4), 347-359.

Heilbrun, K., Douglas, K. S., & Yasuhara, K. (2009). Violence risk assessment: Core controversies. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy* (pp. 333-357). New York, NY: Guilford.

Heilbrun, K., Nezu, C. M., Keeney, M., Chung, S., & Wasserman, A. L. (1998). Sexual offending: Linking assessment, intervention, and decision making. *Psychology, Public Policy, and Law, 4*(1), 138-174.

Hildebrand, M., & de Ruiter, C. (2012). Psychopathic traits and change indicators of dynamic risk factors during inpatient forensic psychiatric treatment. *International Journal of Law and Psychiatry, 35*, 276-288.

Holliday, S. B., Heilbrun, K., & Fretz, R. (2012). Examining improvements in criminogenic needs: The risk reduction potential of a structured re-entry program. *Behavioral Sciences and the Law, 30*, 431-447.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19.

Jones, N. J., Brown, S. L., & Zamble, E. (2010). Predicting criminal recidivism in adult make offenders: researcher versus parole officer assessment of dynamic risk. *Criminal Justice and Behavior, 37*(8), 860-882.

Kazdin, A. E., Kraemer, H. C., Kessler, R. C., Kupfer, D. J., & Offard, D. R. (1997). Contributions of risk-factor research to developmental psychopathology. *Clinical Psychology Review, 17*(4)*,* 375-406.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54,* 337-343.

Kraemer, H. C., Stice, E., Kazdin, A. E., Offord, D. R., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, 158*(6), 848-856.

Landis, J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lavoie, J. A. A., Guy, L. S., & Douglas, K. S. (2009). Violence risk assessment: Principles and models bridging prediction to management. In J. Ireland (Ed.), *Violent and sexual offenders: Assessment, treatment and management* (pp. 3-26). Portland, OR: Willan.

Lewis, K., Olver, M. E., & Wong, S. C. P. (2013). The Violence Risk Scale: Predictive validity and linking changes in risk with violent recidivism in a sample of high-risk offenders with psychopathic traits. *Assessment, 20*(2), 150-164.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13-22.

Lussier, P., & Davies, G. (2011). A person-oriented perspective on sexual offenders, offending trajectories, and risk of recidivism: A new challenge for policymakers, risk assessors, and actuarial prediction. *Psychology, Public policy, and Law, 17*(4), 530-561.

McGlinchey, J. B., Atkins, D. C., & Jacobson, N. S. (2002). Clinical significant methods: Which one to use and how useful are they? *Behavior Therapy, 33*, 529-550.

McGlinchey, J. B., & Jacobson, N. S. (1999). Clinically significant but impractical? A response to Hageman and Arrindell. *Behaviour research and Therapy, 37*, 1211-1217.

Michel, S. F., Riaz, M., Webster, C., Hart, S. D., Levander, S., Muller-Isberner, R., Tiihonen, J., et al. (2013). Using the HCR-20 to predict aggressive behaviour among men with schizophrenia living in the community: Accuracy of prediction, general and forensic settings, and dynamic risk factors. *International Journal of Forensic Mental Health, 12*(1), 1-13.

Monahan, J. (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice & Behavior, 23*(1)*,* 107-120.

Monahan, J., & Steadman, H. J. (2001). Violence risk assessment: A quarter century of research. In L. E. Frost & R. J. Bonnie (Eds.), *The evolution of mental health law* (pp. 195-211). Washington, DC: American Psychological Association.

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., Roth, L. H., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of mental disorder and violence.* New York, NY: Oxford University Press.

Mossman, D. (1994). Assessing predictions about violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.

Mulvey, E. P., & Lidz, C. W. (1995). Conditional prediction: A model for research on dangerousness to others in a new era. *International Journal of Law & Psychiatry, 18*(2)*,* 129-143.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*(5)*,* 1181-1196.

Neves, A. C., Goncalves, R. A, & Palma-Oliveira, J. M. (2010, May). *Dynamic risk assessment: An analysis of change with the HCR-20*. Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vancouver, Canada.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*(3), 328-362.

Nicholls, T. L., Brink, J., Desmarais, S. L., Webster, C. D., Martin, M. (2006). The Short-Term Assessment of Risk and Treatability (START): A prospective validation study in a forensic psychiatric sample. *Assessment, 13*(3), 313-327.

Nicholls, T. L., Viljoen, J. L., Cruise, K. R., Desmarais, S. L., & Webster, C. D. (2010). *Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV) (abbreviated manual)*. Coquitlam, Canada: British Columbia Mental Health and Addiction Services.

Nonstad, K., Nesset, M. B., Kroppan, E., Pedersen, T. W., Nottestad, J. A., Almvik, R., & Palmstierna, T. (2010). Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian high secure hospital. *International Journal of Forensic Mental Health, 9*(4), 294-299.

Odgers, C. L., Mulvey, E. P., Skeem, J. L., Gardner, W., Lidz, C. W., & Schubert, C. (2009). Capturing the ebb and flow of psychiatric symptoms with dynamical systems models. *American Journal of Psychiatry, 166*(5), 575-582.

Olver, M. E., Lewis, K., & Wong, S. C. P. (2013). Risk reduction treatment of high-risk psychopathic offenders: The relationship of psychopathy and treatment change to violent recidivism. *Personality Disorders: Theory, Research, and Treatment, 4*(2), 160-167.

Olver, M. E., & Wong, S. C. P. (2011). A comparison of static and dynamic assessment of sexual offender risk and need in a treatment context. *Criminal Justice and Behavior, 38*(2), 113-126.

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). 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.

Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology, 56*(10), 1239-1262.

Otto, R. K., & Douglas, K. S. (Eds.). (2010). *Handbook of violence risk assessment*. New York, NY: Routledge/Taylor & Francis Group.

Pell, T., & Hargreaves, L. (2011). Making sense of cluster analysis: Revelations from Pakistani science classes. *Cambridge Journal of Education, 41*(3), 347-367.

Prochaska, J. O., & DiClemente, C. C. (1986). Toward a comprehensive model of change. In W. Miller & N. Heather (Eds.), *Treating addictive behaviors* (pp. 3-27). New York, NY: Plenum Press.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Quinsey, V. L., Jones, B. G., Book, A. S., & Barr, K. N. (2006). The dynamic prediction of antisocial behavior among forensic psychiatric patients: A prospective field study. *Journal of Interpersonal Violence, 21*(12), 1539-1565.

Raynor, P. (2007). Risk and need assessment in British probation: The contribution of the LSI-R. *Psychology, Crime, and Law, 13*(2), 125-138.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

Salekin, R.T. (2002). Psychopathy and therapeutic pessimism: Clinical lore or clinical reality? *Clinical Psychology Review, 22*, 79-112.

Schlager, M. D., & Pacheco, D. (2011). An examination of changes in LSI-R scores over time: Making the case for needs-based case management. *Criminal Justice and Behavior, 38*(6), 541-553.

Shah, S. A. (1978). Dangerousness: A paradigm for exploring some issues in law and psychology. *American Psychologist, 33*, 224-238.

Simourd, D. J. (2004). Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior, 31*(3), 306-323.

Skeem, J. L., Monahan, J., & Mulvey, E. P. (2002). Psychopathy, treatment involvement, and subsequent violence among civil psychiatric patients. *Law and Human Behavior, 26*(4), 577-603.

Skeem, J. L., & Mulvey, E. (2002). Monitoring the violence potential of mentally disordered offenders being treated in the community. In A. Buchanan (Ed.), *Care of the mentally disordered offender in the community* (pp. 111-142). New York, NY: Oxford Press.

Skeem, J. L., Mulvey, E. P., & Lidz, C. W. (2000). Building mental health professionals' decisional models into tests of predictive validity: The accuracy of contextualized predictions of violence. *Law and Human Behavior, 24*(6), 607-628.

Skeem, J. L., Mulvey, E., Lidz, C., Gardner, W., & Schubert, C. (2002). Identifying psychiatric patients at risk for repeated involvement in violence: The next step toward intensive community treatment programs. *International Journal of Forensic Mental Health Services, 1*(2), 155-170.

Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology, 63*(6), 1044-1048.

Steadman, H. J. (2000). From dangerousness to risk assessment of community violence: Taking stock at the turn of the century. *Journal of the American Academy of Psychiatry and the Law*, *28*, 265-271.

Steadman, H. J., Monahan, J., Robbins, P. C., Appelbaum, P., Grisso, T., Klassen, D., Mulvey, E., & Roth, L. (1993). From dangerousness to risk assessment: Implications for appropriate research strategies. In S. Hodgins (Ed.), *Crime and mental disorder* (pp. 39-62). Thousand Oaks, CA: Sage.

Viljoen, J. L., Beneteau, J. L., Gulbransen, E., Brodersen, E., Desmarais, S. L., Nicholls, T. L., & Cruise, K. R. (2012). Assessment of multiple risk outcomes, strengths, and change with the START:AV: A short-term prospective study with adolescent offenders. *International Journal of Forensic Mental Health, 11*(3), 165-180.

Vincent, G. M., Chapman, J., & Cook, N. E. (2011). Risk-needs assessment in juvenile justice: Predictive validity of the SAVRY, racial differences, and the contribution of needs factors. *Criminal Justice and Behavior, 38*(1), 42-62.

Webster, C. D., Douglas, K. S., Belfrage, H., & Link, B. (2000). Capturing change: An approach to managing violence and improving mental health. In S. Hodgins & R. Muller-Isberner (Eds.), *Violence among the mentally ill* (pp. 119-144). Dordrecht, the Netherlands: Kluwer/Academic.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence* (version 2). Burnaby, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Short-Term Assessment of Risk and treatability (START)* (Version 1.1). Coquitlam, BC, Canada: Forensic psychiatric Services Commission.

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., Hart, S. D., & Brink, J. (2013). Predictive validity of dynamic factors: Assessing violence risk in forensic psychiatric inpatients. *Law and Human Behavior*, advance online preview.

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82*(1), 50-59.

Wong, S., & Gordon, A. E. (1999-2003). *The Violence Risk Scale.* Unpublished manuscript, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law, 12*(3), 279-309.

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

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*(5), 740-767.

# Appendices

# Appendix A.

# Items on the HCR-20 and START Risk Assessment Tools

***Table A1. Items on the HCR-20 Risk Assessment Protocol***

| Subscales | Items |
|---|---|
| Historical Scale | |
| H1 | Previous violence |
| H2 | Young age at first violent incident |
| H3 | Relationship instability |
| H4 | Employment problems |
| H5 | Substance use problems |
| H6 | Major mental illness |
| H7 | Psychopathy |
| H8 | Early Maladjustment |
| H9 | Personality disorder |
| H10 | Prior supervision failure |
| Clinical Scale | |
| C1 | Lack of insight |
| C2 | Negative attitudes |
| C3 | Active symptoms of major mental illness |
| C4 | Impulsivity |
| C5 | Unresponsive to treatment |
| Risk Management Scale | |
| R1 | Plans lack feasibility |
| R2 | Exposure to destabilizers |
| R3 | Lack of personal support |
| R4 | Noncompliance with remediation attempts |
| R5 | Stress |

*Note*. Adapted from Webster et al. (1997).

***Table A2. Items on the START Risk Assessment Protocol***

| | Items |
|---|---|
| 1 | Social Skills |
| 2 | Relationships |
| 3 | Occupational |
| 4 | Recreational |
| 5 | Self-Care |
| 6 | Mental State |
| 7 | Emotional State |
| 8 | Substance Use |
| 9 | Impulse Control |
| 10 | External Triggers |
| 11 | Social Support |
| 12 | Material Resources |
| 13 | Attitudes |
| 14 | Medication Adherence |
| 15 | Rule Adherence |
| 16 | Conduct |
| 17 | Insight |
| 18 | Plans |
| 19 | Coping |
| 20 | Treatability |

*Note*. Adapted from Webster et al. (2009).