Risk State and Purportedly Dynamic Risk Factors: An Examination of Change in Violence Risk Across Multiple Assessments

by

Adam J. E. Blanchard

Master of Arts, Simon Fraser University, 2013

Bachelor of Arts (Hons.), Simon Fraser University, 2010

Associate of Arts, Capilano University, 2007

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Psychology
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY

Fall 2021

Declaration of Committee

Name:	**Adam J. E. Blanchard**
Degree:	**Doctor of Philosophy (Psychology)**
Thesis title:	**Risk State and Purportedly Dynamic Risk Factors: An Examination of Change in Violence Risk Across Multiple Assessments**
Committee:	**Chair:** Ralph Mistlberger Professor, Psychology
	Kevin S. Douglas Supervisor Professor, Psychology
	Stephen D. Hart Committee Member Professor, Psychology
	Kate L. Slaney Committee Member Professor, Psychology
	Maaike Helmus Examiner Assistant Professor, Criminology
	Michael Daffern External Examiner Professor, Clinical Forensic Psychology Swinburne University of Technology

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.

A copy of the approval letter has been filed with 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 2016

Abstract

Professionals have increasingly stressed the dynamic nature of violence risk. To be effective at preventing violence, it is essential to reliably and validly measure dynamic risk factors that can alter an individual's risk state. It is imperative to understand the changeability of dynamic risk factors and global ratings of risk state more completely. There is limited, but promising, research to date that has investigated the ability of dynamic factors on structured risk assessment instruments to change over time (internal responsiveness) and even less research has investigated the extent to which this change is associated with violence (external responsiveness). The current study aimed to add to the existing knowledge concerning the responsiveness of dynamic ratings using the HCR-20 and START. With a combined sample of offenders and civil psychiatric patients, participants were assessed approximately monthly for up to six assessments. The first research question addressed internal responsiveness. Although most ratings did not change across assessments, intraindividual change was not uncommon. It was more common for participants to change at least once across all six assessments on nearly all ratings. Aggregate-level change was also observed for most dynamic ratings on both instruments. In contrast to raw score change, reliable change was relatively rare. The second research question addressed external responsiveness. Many of the dynamic ratings were predictive of violence across assessments when controlling for the respective baseline scores, and when controlling for the Historical scale scores. In contrast, when change scores were analyzed, very few items were predictive of violence. Overall, the current study added to the growing body of empirical research supporting the responsiveness of dynamic ratings on the HCR-20 and START. As such, there is accumulating evidence that supports these items as meeting all the criteria for dynamic risk factors. Repeated assessments of dynamic risk factors and risk state aid in identifying targeted management strategies, monitoring the effectiveness of these strategies, and altering the intensity or target of these strategies.

Keywords: violence risk assessment; dynamic risk factors; risk status and risk state; change over time; HCR-20; START

I dedicate my dissertation to Zero and Duke for their tireless support and aggravation.

Acknowledgements

I would like to first express my deepest gratitude to my supervisor, Dr. Kevin Douglas, for his amazing mentorship, support, and inspiration over the past decade. I am truly grateful for the innumerable opportunities he has provided me. Beginning as an undergraduate student through the completion of two graduate degrees, his mentorship has been vital to my academic and professional development. I would also like to thank my committee members and mentors, Dr. Kathleen Slaney and Dr. Stephen Hart. They have both been magnificent instructors, supervisors, and mentors throughout my education. I am extremely appreciative of all their support over the years.

I also gratefully acknowledge the efforts of all the members of the Douglas Lab over the years who assisted on the two projects that provided the data used here. This dissertation would not have been possible without all the time and dedication they invested over the years. I would like to give a special thanks to Michelle Collins for her painstaking efforts and excellent management of the lab. Her contribution to this project and the mentorship of students in the Douglas Lab is incalculable.

I would also like to thank my parents for their ongoing love and support. Their patience and guidance have been indispensable. Without them, I could not have achieved this level of success. Finally, I would like to thank my love and partner in crime, Glenys Thorp. She has been an amazing support, inspiration, and partner over the years.

Table of Contents

List of Tables

List of Acronyms

AUC	Area under the curve from receiver operation characteristics
GEE	Generalized estimating equations
HCR-20	Historical-Clinical-Risk Management-20
ICC	Intraclass correlation coefficient
MDC	Minimally detectable change
OR	Odds ratio
RCI	Reliable change index
SEM	Standard error of measurement
START	Short-Term Assessment of Risk and Treatability

Chapter 1. Introduction

The complex and consequential charge of assessing and managing an individual's risk of violence is a regular undertaking by a diverse array of professionals across countless settings (e.g., Douglas & Kropp, 2002; Douglas & Otto, 2020, Durrant, 2016; Lyon et al., 2001; Melton et al., 2007; Murrie & Kelley, 2017; Shah, 1978; Singh et al., 2011, 2014). Professionals in innumerable contexts and settings, coming from diverse educational and practice backgrounds including law enforcement, correctional, and health care personnel, are responsible for the challenging task of assessing the likelihood that a given person will engage in a specific type of harm over a certain period of time and attempting to mitigate this risk of harm.

The assessment of an individual's risk of future violence is a formidable task with potentially dire, life-altering consequences for both the person being assessed, as well as the assessors themselves. Depending on the context, the professional opinions resulting from a violence risk assessment may be a factor that leads to momentous impacts on the individual being assessed, ranging from involuntary civil commitment or denial of release to mandated treatment or imposed sanctions; these assessments may result in the individual forfeiting a host of well-established civil liberties (Barefoot v. Estelle, 1983; Canadian Charter of Rights and Freedoms, 1982; Cording et al., 2016; Lyons v. the Queen, 1987; Monahan et al., 2001; United States v. Salerno, 1987). Alternatively, failure to identify an individual at high risk of engaging in violence may result in another individual or the general public being placed in danger, and ultimately result in the safety, rights, or freedoms of another individual being removed (Monahan, 1993; Murrie & Kelley, 2017; Schlesinger, 1996). Moreover, the failure to conduct a comprehensive violence risk assessment or any improper conduct in completion of a risk assessment may have substantial impacts on the professional conducting the assessment, including being open to civil liabilities and licensure or registration forfeiture (Monahan, 1996; Tarasoff v. Regents of California, 1976). As a result of the immense consequence of these assessments, it is imperative that professionals conduct themselves in accordance with the best available practice standards.

The research and professional practice of violence risk assessment has witnessed indispensable and vital advances over the past four decades (e.g., Blanchard

et al., 2016; Cording et al., 2016; de Vogel & de Ruiter, 2006; Douglas & Kropp, 2002; Douglas & Ogloff, 2003; Douglas & Otto, 2020; Hanson, 2005; Heffernan & Ward, 2017; Lavoie et al., 2009; McNeil, 1998; Mills, 2005, 2017; Monahan, 1996; Monahan & Skeem, 2016; Monahan & Steadman, 2001; Steadman, 2000). Early research and literature were replete with skepticism and debate regarding a professional's ability to accurately forecast violence (e.g., Bonta, 1996; Ennis & Litwack, 1974; Monahan, 1981; Steadman & Cocozza, 1974). Since this initial research and scholarship reporting on the ineffectuality of predictions of dangerousness, the field has seen thousands of independent evaluations and dozens of meta-analyses supporting the overall reliability and predictive validity of various structured risk assessment approaches and instruments (e.g., Campbell et al., 2009; Fazel et al., 2012; Guy, 2008; Hanson & Morton-Bourgon, 2009; Singh et al., 2011, 2014; van der Put et al., 2019; Yang et al., 2010).

Throughout the previous decades, the field has seen extensive progress in appreciating the complex and multifaceted nature of violence, identifying and validating evidence-based risk factors for violence, as well as developing and evaluating risk assessment models and specific tools from these models (Andrews & Bonta, 2010b; Bonta & Andrews, 2017; Douglas & Kropp, 2002; Douglas & Skeem, 2005; Durrant, 2016; Harris et al., 2015; Klepfisz et al., 2016; Mills, 2017; Monahan et al., 2001; Quinsey, Harris, et al., 2006). A particularly significant development over the past several decades has been the increased attention to the dynamic or fluctuating nature of violence risk (Douglas & Skeem, 2005; Klepfisz et al., 2016; Lussier & Davies, 2011; Olver & Stockdale, 2020; Serin et al., 2016; Ward, 2015, 2016). Violence risk is considered, in theory, inherently dynamic; criminal justice and mental health systems are founded on the principle that an individual's risk can be controlled and minimized (Olver & Stockdale, 2020; Olver & Wong, 2011).

There is a pressing need to employ empirically supported strategies for assessing and managing risk for violence in forensic, correctional, and inpatient contexts (Dexter & Vitacco, 2020; Martin & Daffern, 2006; McDermott et al., 2008; Vitacco et al., 2012). There is currently a large number of established, validated instruments designed to assess the risk of different types of violence (Campbell et al., 2009; Douglas & Otto, 2020; Singh et al., 2014). As such, professionals are tasked with the intricate decision of selecting the most appropriate instrument to use in a given context (Coupland & Olver, 2018). Assessing an individual's risk for violence is one of the most imperative and

consequential functions performed by professionals in diverse contexts. However, it is equally imperative to be able to accurately detect fluctuations in an individual's risk to effectively manage this risk (Brown et al., 2009). It is necessary that contemporary tools for violence risk assessment allow for accurate and reliable identification of dynamic risk factors that can then be targeted with management strategies (Caudy et al., 2013). Currently, there remains various opportunities for refining and enhancing the composition and practice of violence risk assessments, especially surrounding the dynamic nature of risk (Durrant, 2016).

Numerous scholars originating from distinct theoretical and practical contexts have highlighted the dynamic nature of risk factors and general risk level (e.g., Andrews & Bonta, 2010b; Douglas & Skeem, 2005; Hanson & Harris, 2000; Olver & Stockdale, 2020; Serin et al., 2016, Ward, 2016). Now, more than ever, the field stresses that an individual's risk for violence and many of the factors contributing to this risk are theoretically capable of changing over time and context. A major emphasis now centers on a more complete and detailed understanding of the changeable features of violence risk and violence risk factors.

To fully elucidate the dynamic nature of violence risk, it is essential to review several important theoretical developments in the field. To begin with, the ultimate purpose and characteristics of a comprehensive violence risk assessment will highlight the inherent connection between assessment and management of risk. Next, the distinction between different types of risk factors and risk markers will be reviewed, focusing on their purported ability to change over time (i.e., static versus dynamic risk factors). The importance of considering the dynamic or changeable aspects of risk and risk factors will also be examined for both the prediction and management of violence. As well, it is imperative to detail underlying frameworks relevant to measuring change in violence risk assessment ratings using various psychometric properties. Subsequently, the available empirical evidence relevant to the notion of dynamic risk will be examined with a focus on studies that have examined ratings on two commonly used structured violence risk assessment tools.

Over a decade ago, it was noted that the "exploration of the predictive validities of assessments of change remains a major issue and is, perhaps, the major issue for the development of theory and practice in the psychology of crime" (Andrews, 1995, p. 54,

as cited in Brown et al., 2009). Although considerable theoretical and a growing body of empirical evidence has been added to our understanding of dynamic risk, it is still the case that “the science of risk assessment currently lags far behind practice” (Douglas & Skeem, 2005, p. 367). That is, the accurate identification and “assessment of dynamic risk factors is a major concern of contemporary researchers and practitioners” (Heffernan & Ward, 2015, p. 250; see also Mann et al., 2010; Ward, 2014). Moreover, “identifying dynamic risk factors that are related to recidivism is a pressing challenge in the field” (Caudy et al., 2013, p. 458). A thorough review of the dynamic risk literature demonstrates that the extensive body of theoretical developments and knowledge currently dwarfs the scarcity of empirical investigations (Douglas & Skeem, 2005; Monahan & Skeem, 2016; Olver & Stockdale, 2020).

As a result, the current research was undertaken in attempt to add to the growing body of empirical investigations into dynamic risk using contemporary structured violence risk assessment instruments. This research investigated change in overall risk (state) and change in theoretically dynamic risk factors on two prominent, structured violence risk assessment tools. More precisely, the current study sought to identify risk factors capable of intraindividual change on these risk assessment instruments, as well as whether global assessments of violence risk made using these instruments evidenced intraindividual change. In addition, the study sought to describe some of the characteristics of change in risk factors and global risk level. Finally, the study aimed to determine whether change in risk level or risk factors was associated with similar changes in the perpetration of future violence.

1.1. Comprehensive Violence Risk Assessments

The theory and practice of violence risk assessment have been the focus of extensive literature for several decades (e.g., Heilbrun, 1997; Kozol et al., 1972; Litwack & Schlesinger, 1999; Monahan, 1981, 1984, 1996; Scott, 1977; Shah, 1978; Tardiff, 1996; Webster & Menzies, 1989). Throughout this time, the field has seen extensive growth from the formerly narrow and unsophisticated prediction of dangerousness to the broad and complex focus on a comprehensive violence risk paradigm (Douglas & Kropp, 2002; Dvoskin & Steadman, 1994; Haggård-Grann, 2007; McNiel et al., 2002; Monahan, 1996; Skeem et al., 2000). A significant development has been the unification of the risk assessment process and the risk management process (i.e., prevention and treatment;

Douglas & Kropp, 2002; Douglas & Skeem, 2005; Hart, 1998; Heilbrun, 1997). Over the course of this growth, the ultimate focus or purpose of most violence risk assessments has shifted from the one-time prediction of violence to the ongoing appraisal of risk as a process to identify and implement strategies to minimize or eliminate this risk (Douglas, Cox, et al., 1999; Douglas & Skeem, 2005; Dvoskin & Heilbrun, 2001; Dvoskin et al., 2012; Heilbrun, 1997; Hart, 1998; McNiel et al., 2002; Otto, 2000; Skeem et al., 2000; Steadman, 2000; Steadman et al., 1993; Webster et al., 2000). Thus, the field has seen a noteworthy shift in that the fundamental aim of a comprehensive risk assessment has become the prevention, not prediction, of violence (Douglas & Kropp, 2002; Hart, 1998; Heilbrun, 1997).

Currently, a comprehensive violence risk assessment embodies two distinct, yet interrelated, stages (Douglas, Blanchard, et al., 2013; Guy et al., 2015; Heilbrun, 1997; Heilbrun et al., 1998; McNiel et al., 2002; Monahan & Skeem, 2016). The first stage is focused on identifying and understanding the individual's potential for perpetrating violence in the future (Douglas, Blanchard, et al., 2013; Guy et al., 2015; Heilbrun, 1997). This stage characteristically involves the systematic examination of the individual's past and present psychological and social functioning as well as their history of violence and other antisocial behaviours. Conventionally referred to as the process of *prediction*, it was largely about making single time-point estimations of the likelihood of violence (McNiel et al., 2002). Presently, this process differs from producing probabilistic estimates of violence (i.e., exact quantitative assertions of the probability of perpetrating violence over a fixed period of time), as the chief goal is achieving an understanding of the individual's potential (i.e., what this individual is capable of doing and might do in the future; Guy et al., 2015). The ultimate aim of this stage is the development of an understanding regarding various facets of the individual's risk for violence, including the type, severity, frequency, likelihood, and potential victims of violence the individual may perpetrated in the future, as well as the specific contexts in which the violence may occur, and the underlying reasons contributing to this individual's risk of violence.

The second stage is focused on identifying and understanding the factors that might mitigate, moderate, or eliminate the individual's potential for perpetrating violence (Douglas, Blanchard, et al., 2013; Guy et al., 2015; Heilbrun, 1997). This stage characteristically involves the systematic examination of an extensive array of features related to the individual's present and future circumstances, including personal, social,

and contextual factors, so as to develop management strategies targeted to mitigate or alleviate the individual's risk of violence. Conventionally referred to as the process of *risk management*, it is largely about conducting assessments that direct intervention and management strategies to prevent violence. The ultimate aim of this stage is the prevention of violence or the reduction of harm caused by any violence. That is, the focus in this stage is on lessening the individual's risk of violence by identifying and implementing individualized management strategies. As such, the management stage highlights the direct relationship between assessment and management, as a comprehensive assessment should direct the intensity, type, and target of each management strategy. Furthermore, this stage stresses the importance of ongoing, repeated assessments with a focus on dynamic risk factors, in particular those factors that are modifiable through targeted intervention strategies.

There is a consensus in the field that the fundamental aim of each of these stages is noticeably different, yet debate has been present for years regarding whether each of the stages can be conducted independently (Douglas, Blanchard, et al., 2013; Guy et al., 2015; Harris et al., 2015; Heilbrun, 1997; McNiel et al., 2002; Mills, 2017). According to some professionals, these two stages comprise distinct models that are entirely distinguishable and independent (e.g., Harris et al., 2015; Heilbrun, 1997; Scurich, 2016; Ward & Beech, 2015). Proponents of this paradigm claim there is a simple, temporal relationship between the two stages; professionals begin with the prediction stage and then, only if necessary, continue on to the management stage. Thus, the prediction stage can be conducted independent of and without consideration to the management stage. According to supporters of this view, certain criminal justice contexts require only the prediction of violence (i.e., the relevant psycholegal issue concerns exclusively the probability that an individual will engage in future violence), whereas other criminal justice and clinical contexts require consideration of both the prediction and management of violence.

According to other professionals, the prediction and management stages are indivisible and inter-reliant (e.g., Cooke & Michie, 2013; Guy et al., 2015; Hart et al., 2001; McNiel et al., 2002; Mills, 2017; Murrie & Kelley, 2017). Proponents of this paradigm claim that the relationship between the two stages is "bidirectional or recursive" (Guy et al., 2015, p. 57); professionals begin with the prediction stage and then inevitably continue on to the management stage, which may result in reiteration of

and modifications to the prediction stage and consequently reiteration of and modifications to the management stage. Thus, judgments arrived at during one stage will impact judgments made at the other and vice versa; the two stages are inseparable. According to supporters of this view, professionally responsible and ethical prediction is not achievable without a systematic consideration of possible management strategies and the individual's future circumstances. As well, professionally responsible and ethical risk management must be based on the systematic consideration of the individual's potential for future violence. This paradigm (i.e., a comprehensive risk assessment involves both inseparable stages) now dominates the field.

As a result of multiple changes (e.g., legal, clinical, societal, and economic), the fundamental aim of comprehensive risk assessments now centers on the ongoing assessment of risk in order to develop and implement strategies to reduce or prevent violence (van den Berg et al., 2020; Dexter & Vitacco, 2020; Douglas, Cox, et al., 1999; Douglas & Skeem, 2005; Dvoskin & Heilbrun, 2001; Heilbrun, 1997; Hart, 1998; Otto, 2000; Skeem et al., 2000; Steadman, 2000; Steadman et al., 1993; Webster et al., 2000). As such, the ultimate purpose of a comprehensive violence risk assessment is "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 Douglas & Reeves, 2010; Hart, 2008; McNiel et al., 2002). That is, a comprehensive violence risk assessment is focused on "violence *prevention* more than violence *prediction*, the aim is not only an overall risk estimate but, rather, ongoing identification and mitigation of any factors that may be conducive to violence" (Murrie & Kelley, 2017, p. 135, emphasis in original). In other words, the aim of a comprehensive risk assessment is to "detect the relevant phenomena associated with [violence], infer the mechanisms that generate them, and to then construct an intervention plan that seeks to modify or remove the causes" (Ward, 2016, p. 6). Accordingly, a comprehensive violence risk assessment necessitates both the prediction and management stages (Mills, 2017). Moreover, a comprehensive risk assessment must incorporate the consideration of dynamic risk factors that can serve as intervention targets and allow for identifiable changes in the individual's potential for future violence (Dexter & Vitacco, 2020; Ward, 2016).

1.2. Types of Risk and Risk Factors

It has been recognized for many years that "the absence of precise language is perhaps the major problem in current risk research" (Kraemer, 2003, p. 41). This is undoubtedly true in the context of violence risk assessment research and practice. That is, unambiguous and detailed conceptualizations and terminology are required in order to progress the current state of the violence risk assessment field (Monahan & Skeem, 2016). Both theoretical and empirical developments in the field benefit from the use of well-defined notions of risk for violence and risk factors for violence. As such, it is necessary to clearly articulate the main types of risk and various typologies of risk factors currently in use throughout the field.

1.2.1. Risk State and Risk Status

Under the framework of comprehensive violence risk assessments involving both the assessment and management of violence risk, there exists a crucial distinction between the concepts of risk status and risk state (see Table 1-1; Douglas & Skeem, 2005; Heilbrun et al., 2009; Klepfisz et al., 2016; Monahan, & Skeem, 2016; Skeem & Mulvey, 2002). In this context, *risk status* denotes interindividual differences in risk level. That is, risk status refers to identifiable differences in global risk level between different groups of individuals; some individuals present an increased risk compared to other individuals based on their elevated status on known risk factors for violence. Risk status is based largely on the individuals' standing on static (i.e., unchangeable) risk factors; as such, it is considered principally invariant over time. An individual's risk status provides a baseline level of risk and dictates the required level or intensity of intervention required to mitigate that risk (Klepfisz et al., 2016).

In contrast, *risk state* denotes the intraindividual differences in risk level posed by a particular individual at a given moment in time (Douglas & Skeem, 2005; Heilbrun et al., 2009; Klepfisz et al., 2016; Monahan, & Skeem, 2016; Skeem & Mulvey, 2002). Amongst those identified as high, moderate, or low risk based on risk status, a given individual's risk of future violence inevitably "ebbs and flows" across time and circumstance (Douglas & Skeem, 2005, p. 348). Risk state refers to a given individual's proclivity for perpetrating violence at a specified moment in time based on their current standing on biological, psychological, social, and contextual factors. As a given

individual experiences fluctuations and modifications to these various domains, they similarly experience fluctuations in the risk they pose for future violence. Risk state is based on the individual's current standing on the combination of static (i.e., unchangeable) and dynamic (i.e., changeable) risk factors with a greater emphasis on dynamic risk factors; accordingly, it is considered principally variable over time and circumstance. An individual's risk state is theorized to vary across time due to fluctuations in both external (e.g., treatment, supervision, monitoring) and internal (e.g., learning, aging, motivation) forces.

The distinction between risk status and risk state is comparatively new to the field of violence risk assessment (Douglas & Skeem, 2005; Skeem & Mulvey, 2002). Nevertheless, this distinction and the importance of considering an individual's risk state has gained extensive acceptance. Most scholars and practitioners consider that each individual's risk of perpetrating violence waxes and wanes over time and across contexts in direct relation to the individual's current standing on numerous recognized risk factors (Heilbrun et al., 2009; Monahan, & Skeem, 2016). This understanding accords greatly with the shift in ultimate focal point from prediction to prevention of violence (Douglas & Kropp, 2002; Douglas & Skeem, 2005; Heilbrun, 1997). The belief that it is possible to assess the likelihood an individual will perpetrate violence (i.e., prediction) in order to develop and implement strategies to mitigate this risk (i.e., management) hinges on the belief that risk (state) is capable of fluctuating and being altered by external forces.

1.2.2. Types of Risk Factors

Generally, a *risk factor* is defined as an identifiable characteristic (e.g., biological, psychological, social, or situational) of an individual that (a) precedes the outcome of interest (e.g., violence) and (b) is statistically associated with the outcome (Hart, 2008; Kazdin et al., 1997; Kraemer, 2003; Kraemer et al., 1997). That is, a risk factor is present prior to the perpetration of violence and its presence corresponds with an increase in the likelihood of perpetrating violence compared to its absence. Notably, this definition does not rely on or imply that the risk factor and outcome are causally related (Skeem & Monahan, 2011).

Table 1-1. Risk and Risk Factor Taxonomies

Descriptor	Definition	Example	Source
Risk status	Inter-individual differences in risk based largely on static risk factors; largely time-invariant	-	Douglas & Skeem (2005)
Risk state	Intra-individual differences in risk based largely on current status of dynamic risk factors; time-variant	-	
Risk factor	Any characteristic that precedes the outcome and is statistically related to the outcome	Prior violence	Kraemer et al. (1997) Kraemer (2003)
Fixed risk marker	A risk factor that cannot be demonstrated to change either spontaneously or through intervention	Biological sex	
Variable risk marker	A risk factor that is changeable but not through intervention	Current age	
Variable risk factor	A risk factor that is changeable through intervention	Employment status	
Causal risk factor	A risk factor that is changeable through intervention and when changed corresponds with change in risk of the outcome	Antisocial cognitions	
Static risk factor	A risk factor that cannot be demonstrated to change either spontaneously or through intervention	Childhood abuse	Bonta (1996) Hanson (1998) Heilbrun (1997)
Dynamic risk factor	A risk factor that is changeable either spontaneously or through intervention	Impulsiveness	
Dynamic risk factor / Criminogenic need	A risk factor that is changeable and when changed corresponds with change in the likelihood of the outcome	Antisocial attitudes	Andrews & Bonta (2010b) Douglas & Skeem (2005)
Stable dynamic risk factor	A dynamic risk factor that is a relatively enduring characteristics (i.e., changes over months or years)	Substance abuse problems	Hanson (1998) Hanson & Harris (2000)
Acute dynamic risk factor	A dynamic risk factor that changes more rapidly (i.e., days, hours, or minutes) and may immediately precede the outcome	Negative mood	
Psychologically meaningful risk factor	A propensity that (a) has a plausible justification as a cause of offending and (b) predicts offending	Sexual preoccupation	Mann et al. (2010)

Note. Adapted from Durrant (2016); Kraemer (2003); Monahan & Skeem (2014, 2016).

Several notable typologies or classification methods of risk factors have been introduced in the violence risk literature that are relevant to a dynamic conceptualization of risk. These typologies disaggregate risk factors based largely on their ability to fluctuate either spontaneously or through direct manipulation, as well as the influence of these fluctuations on the likelihood of violence. Further methods focus on the degree of association with violence, or relative interaction with other risk factors.

Time Variable and Invariable Risk Factors

Several authors have stressed the importance of clearly articulating the differences in types of risk factors based on their ability to change over time (see Table 1-1). Predominantly, Kraemer and colleagues (1997; see also Dempster & Hart, 2002; Douglas & Kropp, 2002; Douglas & Skeem, 2005; Hart, 2008; Kazdin et al., 1997; Kraemer, 2003) distinguished between various types of risk factors based on their temporal association with the outcome and their ability to change over time.

According to this typology, a correlate is defined as an element or feature of an individual that is statistically associated with the outcome of interest (i.e., its presence increases the likelihood of observing the outcome versus its absence), yet it does not necessarily precede the outcome. That is, there is no established temporal relationship between the correlate and the outcome; it is unknown whether a correlate precedes the outcome or follows it. A risk factor is, thus, a correlate that is shown to precede the outcome of interest. As mentioned, risk factors must both precede and be statistically associated with the outcome of interest.

In this typology, risk factors can be separated into distinct types based on their ability to change over time (fixed versus variable), the manner in which they are capable of change (marker versus factor), and their functional relation to the outcome (causal factor). Fixed factors are time-invariant, whereas variable factors are time-variant. As such, a fixed risk marker is a risk factor that is not capable of change or being changed over time by any means. In contrast, a variable risk marker is a risk factor that is capable of spontaneous change over time, but not capable of manipulation through intervention. A risk factor that is capable of change through direct intervention is called a variable risk factor. If there is no empirical evidence that speaks to the ability of the factor to change over time, then the general term of risk factor is most appropriate. Finally, a risk factor that is capable of change through direct intervention and this change results in a parallel

change on the likelihood of the outcome is referred to as a causal risk factor. However, if the characteristic is capable of change, but this change does not result in a corresponding change to the likelihood of the outcome, then it should be labelled a variable risk marker. On the other hand, if the factor is capable of change through intervention, but no empirical evidence can demonstrate a corresponding change in the likelihood of the outcome, then the term variable risk factor is applied.

Causal risk factors are the most important factors to identify under this typology. These factors must satisfy four conditions; casual risk factors are identifiable characteristics that (a) precede the outcome of interest, (b) are statistically associated with the outcome, (c) are capable of change over time through intervention, and (d) when change occurs on the characteristic, there is a corresponding change in the likelihood of the outcome. Notably, Kraemer and colleagues (1997; Kraemer, 2003) used "the term causal risk factor, not the term cause" (p. 341). Although this framework uses causal language, Kraemer et al. (1997) did not require experimental manipulation to establish the presence of a causal risk factor, as many risk factors can neither be ethically nor reasonably assigned to individuals (Serin et al., 2016). As a result, it is not possible to eliminate the potential that other, unobserved variables are the true driving force behind the measured associations (Serin et al., 2016). Instead, it is necessary to determine a logical connection between the factor and violence, as well as exclude alternative explanations (Haynes, 1992; Klepfisz et al., 2016).

In addition, other scholars have presented similar, yet more concise, typologies of risk factors based on their ability to change (see Table 1-1). These typologies differentiate between static (fixed) risk factors and dynamic (changeable) risk factors (Andrews & Bonta, 2010b; Bonta, 1996; Craissati & Beech, 2003; Douglas & Skeem, 2005; Hanson, 1998; Heilbrun, 1997; Kazdin et al., 1997; Kraemer et al., 1997). Largely, static risk factors are time-invariant, whereas dynamic risk factors are time-variant.

Static risk factors are not capable of change over time or across circumstance; it is not possible for them to change either spontaneously or directly as a result of some intervention or management strategy. Overall, static risk factors are best at indicating an atypical developmental trajectory (e.g., intellectual disability) or an undesirable historical characteristic (e.g., childhood abuse) that indicates a long-term proclivity for engaging in violence (Hanson, 1998). That is, these risk factors speak most to risk status. They are

most useful for separating individuals into groups based on their overall risk level (i.e., based on the presence of a greater number and intensity of static risk factors) and for determining long-term risk for violence (Craig et al., 2005). Although static risk factors can speak to the generally intensity of management strategies required in a given case, they are incapable of monitoring potential changes in risk over time or indicating when violence is more or less likely to occur. Static risk factors can also have a significant bearing on dynamic risk factors, such as childhood attachment patterns influencing later interpersonal relationships.

Dynamic risk factors are capable of change over time and across circumstances; it is possible for them to change either spontaneously or directly as a result of some intervention or management strategy (Craissati & Beech, 2003; Douglas & Skeem, 2005; Hanson, 1998). However, in the context of the assessment and management of violence, there exists ambiguity surrounding the exact definition and conceptualization of a dynamic risk factor (Serin et al., 2016; Ward, 2015, 2016). Depending on the specific conceptualization and source used, there are different criteria for being labeled a dynamic risk factor, and the field has seen continued debate, critique, and reformulation (Douglas & Skeem, 2005; Klepfisz et al., 2016; Serin et al., 2016; Ward, 2015, 2016). According to some typologies (Bonta, 1996; Hanson, 1998; Heilbrun, 1997), the main criteria for a dynamic risk factor is the capacity for change over time. As such, any risk factor that is purportedly capable of change would be labelled as a dynamic risk factor. Under this definition, on the whole, the distinction between static and dynamic risk factors is most closely and more accurately considered as a distinction between fixed risk markers and variable risk factors (Monahan & Skeem, 2014).

On the other hand, Douglas and Skeem (2005) advise that a dynamic risk factor must satisfy four specific criteria; the characteristic must (a) precede violence, (b) increase the propensity for violence, (c) be capable of change over time either spontaneously or as a result of intervention, and (d) be capable of predicting changes in the likelihood of violence when change occurs on the characteristic (see also Klepfisz et al., 2016). Thus, according to Kraemer and colleagues (1997) typology, these characteristics must not only be risk factors, satisfying the first two criteria, but also be variable risk factors, satisfying the third criteria, and causal risk factors, satisfying the fourth criteria.

Similarly, Andrews and Bonta (2010b) referred to changeable risk factors as criminogenic needs. According to these authors, a criminogenic need is a risk factor that is, at least theoretically, capable of change and when changed, there is an associated change in the odds of engaging in criminal activity (or violence). Their description of these criminogenic needs has become synonymous with and equivalent to the concept of dynamic risk factor (Serin et al., 2016). Under this more complex definition, the distinction between static and dynamic risk factors is more accurately considered as a distinction between fixed risk markers and causal risk factors (Monahan & Skeem, 2014). Thus, static risk factors are ultimately considered not useful for directing and monitoring risk reduction strategies, whereas dynamic risk factors, also referred to as criminogenic needs, are not only useful but essential for risk reduction.

The contemporary definition of a dynamic risk factor (e.g., Andrews & Bonta, 2010b; Douglas & Skeem, 2005) is therefore consistent with the definition of a causal risk factor (Kraemer et al., 1997). By definition, dynamic risk factors must be capable of demonstrating change, and change on these factors must be associated with a simultaneous change in the likelihood of perpetrating future violence (Brown et al., 2009; Caudy et al., 2013; Douglas & Skeem, 2005; Hanson & Harris, 2000; Howard & Dixon, 2013; Jones et al., 2010; Lewis et al., 2013; Serin et al., 2016). Therefore, research needs to demonstrate that these purported dynamic risk factors (a) exhibit changes over time and (b) exhibit corresponding changes in the likelihood of engaging in violence (Klepfisz et al., 2016; Serin et al., 2016).

However, “this fact is rarely recognized in current discourse” (Monahan & Skeem, 2016, p. 498). The important distinction between variable risk factors and causal risk factors are often clouded and confused under the headings of dynamic risk factors or criminogenic need. The label dynamic risk factor (or criminogenic need) is often used to reference factors that are theoretically capable of change and alleged to be relevant to risk reduction, yet empirical evaluations have not confirmed these suppositions. Accordingly, many dynamic risk factors should appropriately be considered simply risk factors (i.e., empirical evidence has yet to conclusively demonstrate they are capable of change) or variable risk factors, at best (i.e., empirical evidence has shown they are capable of change but not conclusively demonstrated that the changes result in corresponding changes to the likelihood of violence; Monahan & Skeem, 2016).

Stable and Acute Dynamic Risk Factors

In addition, some scholars have distinguished between different types of dynamic risk factors based on the supposed pace of their ability to change. That is, dynamic risk factors are thought to differ in their rapidity and frequency of change with some changing more quickly compared to others (Douglas & Skeem, 2005; Hanson & Harris, 2000). This has led some to argue that dynamic risk factors can be grouped meaningfully into two categories.

Hanson and Harris (2000; see also van den Berg et al., 2020; Casey, 2016; Craissati & Beech, 2003; Douglas & Skeem, 2005; Hanson, 1998; Hanson et al., 2007) differentiate between stable and acute dynamic risk factors. *Stable dynamic risk factors* can change gradually and over long periods of time (i.e., months or even years) but cannot change rapidly or over short periods of time. These factors are unlikely to change spontaneously, instead requiring effortful processing often involving targeted interventions (van den Berg et al., 2020). These factors are often persistent characteristics, personality features, or enduring traits (Craig et al., 2005); they include learned behaviours, skills, deficits, and self-management styles that are considered impossible to change quickly (van den Berg et al., 2020; Klepfisz et al., 2016).

In contrast, *acute dynamic risk factors* are able to change quickly, frequently, and over short periods of time (i.e., weeks, days, or even hours). These factors can change instantaneously and spontaneously (van den Berg et al., 2020). They often immediately precede the perpetration of violence and, thus, signify a high risk (or increased risk) of imminent violence. As a result, acute factors are argued to have the greatest potential for signalling the time at which violence is most likely to occur (Casey, 2016). Acute dynamic risk factors are infrequently the targets of direct intervention but often important for other risk management strategies (van den Berg et al., 2020; Beech et al., 2002).

Although this theoretical distinction is often cited in the literature, few empirical studies have examined the distinction between different dynamic risk factors based on their speed of change (i.e., between stable and acute factors). As such, there is currently little empirical support for the distinction between stable and acute dynamic risk factors (Mann et al., 2010) and some investigations have called this distinction into question (Thornton, 2016). Moreover, attempts to classify risk factors according to their relative time-variance or speed in ability to change are potentially misleading and oversimplified.

There remains considerable cloudiness and ambiguity between stable and acute dynamic risk factors. Even though the theoretical distinction between static risk factor (i.e., unchangeable) and stable dynamic risk factor (i.e., gradually changeable over extended periods) appears exact and concrete, in reality it is less clear (Craig et al., 2005). For example, psychopathy is most often conceptualized as a static risk factor; it is considered a pervasive and persistent set of personality features and characteristics that are stable over the life span (Hare, 1991). On the other hand, empirical evaluations of the treatment of psychopathy have shown that it can gradually respond to some types of interventions and specific intervention strategies (e.g., Edens et al., 2009; Salekin, 2002; Skeem et al., 2002). From these studies it is apparent that levels of psychopathic traits can fluctuate over time. More broadly, meta-analytic studies have shown that personality traits can change gradually over time due to aging (Roberts et al., 2006) and change more rapidly due to clinical interventions (Roberts et al., 2017). As such, some scholars argue that the changeability of risk factors should be viewed along a continuum, extending from completely fixed risk markers or static risk factors (e.g., age at first offence) to extremely variable or acute dynamic risk factors (e.g., intoxication; Brown et al., 2009; Hanson & Harris, 2000; Quinsey, Harris, et al., 2006).

Psychologically Meaningful Risk Factors

Instead of classifying risk factors based on their relative ability to change over time, another conceptualization introduces the concept of psychologically meaningful risk factors (see Table 1-1). According to Mann and colleagues (2010), a *psychological meaningful risk factor* is an individual propensity that may or may not materialize during any given timeframe. Propensities, similar to traits, comprise enduring characteristics of the individual (i.e., emotions, cognitions, and behaviours). These features are regarded as propensities to stress that the outcome of interest (i.e., violence) occurs through interactions between the individual and their environment (Mann et al., 2010; Casey, 2016). Within the context of comprehensive violence risk assessments aimed at preventing violence, “the most useful propensities are those that are amenable to change” (Mann et al., 2010, p. 195). However, under this framework, the ability to change is not a required feature of psychologically meaningful risk factors and practitioners should not solely focus on the amenability to change through intervention for the risk factor to be relevant to treatment.

Proponents of this framework assert that both theory and empirical evidence are relevant to identifying psychologically meaningful risk factors (Mann et al., 2010). Applied originally to sexual offending, the primary consideration is the existence of a reasonable explanation for how the risk factor might be a cause of the outcome of interest (e.g., sexual offending or violence in general). In order to establish that a credible justification exists, there should be a general consensus that the risk factor "(a) is psychologically meaningful, (b) could plausibly be a cause of sexual offending, (c) might be worth targeting in treatment or is already usually targeted in treatment, or (d) is treated as plausible in criminological or social learning theories of offending" (Mann et al., 2010, p. 195). In addition, there should be sound and conclusive empirical support that the propensity does in fact predict the outcome. Finally, additional evidence is necessary in order to firmly establish that a given factor is a cause of violence. Although the authors note that there is no solitary methodology for identifying causal connections, one form of evidence encompasses empirical observations of change in the outcome following experimental manipulation of the factor.

This framework highlights the role and importance of environmental interactions (Casey, 2016). Individual propensities or risk factors are difficult to judge in isolation. Psychologically meaningful risk factors as individual propensities stress the critical influence of context on the manifestations of individual proclivities. This model draws attention to the function and significance interactions play in the etiology of violence.

1.2.3. Interplay of Multiple Risk Factors

In addition to the different types of risk factors and their individual functions in violence risk assessment and management, it is equally imperative to consider the interrelations and interactions between different risk factors (Kraemer, 2003; Kraemer et al., 1997, 2001). The interrelatedness of multiple risk factors and their interactions may play a unique role in an individual's global level of risk for violence. Violence is multifaceted and multidetermined; no single cause or risk factor has been identified or is expected to account for all violence. Multiple causal chains operate in parallel involving various genetic, biological, psychological, social, and environmental risk factors. As such, the "effect of no one of these risk factors can be fully understood except in the context of all the others" (Kraemer et al., 2001, p. 848).

It has been long theorized and acknowledged that “the relations among risk factors and violence can be direct or indirect (mediated), unidirectional or bidirectional, or interactive (moderated)” (Douglas & Skeem, 2005, p. 368). In considering the interrelations amongst a set of risk factors, Kraemer and colleagues (2001; Kraemer, 2003) identified three features that should be considered in order to understand the effects of multiple risk factors on a single outcome. These features are temporal precedence (i.e., which risk factor precedes the others), correlation (i.e., which risk factors are statistically associated with one another), and dominance (i.e., which risk factors alone or in specific combinations leads to the best predictive power).

The combination of these features is considered important in the identification of and prioritization between different causal risk factors for violence (Douglas & Skeem, 2005). By analyzing each of these features in a set of risk factors, it is possible to determine which risk factors are proxy risk factors, overlapping risk factors, independent risk factors, and moderated or mediated by other factors. Although considerable effort has been placed on identifying empirically supported risk factors for violence and differentiating between these various risk factors based on the typologies described above, there is currently a lack of theoretical or empirical knowledge about their interrelationships (van den Berg et al., 2020).

1.3. The Importance of Dynamic Risk

It has been acknowledged for some time that individuals who perpetrate violence are capable of and do in fact change for a variety of underlying reasons and some are able to alter their lives on to better, more prosocial paths without the perpetration of violence (Douglas & Skeem, 2005; Olver & Stockdale, 2020). If the criminal justice and mental health systems are to be successful in their goals of reducing and preventing violence (and crime in general), then it is fundamental to be able to measure reliably and validly those dynamic risk factors that can increase or decrease an individual’s risk state (Beech et al., 2016; Olver & Stockdale, 2020; Olver & Wong, 2011). The focus on violence prevention through changing an individual’s risk state is apparent in the risk assessment tools developed to aid in the decision-making process (Monahan & Skeem, 2016). Most of the initially developed violence risk assessment instruments were designed to assess risk status (i.e., single time-point estimates of the likelihood of engaging in violence based on static risk factors). However, most of the contemporary

decision support tools and their underlying theoretical models focus on risk state with an emphasis on dynamic risk factors (e.g., Andrews & Bonta, 2010b; Douglas & Skeem, 2005; Grann et al., 2005; Webster et al., 1997; Hart, 2008; Monahan, & Skeem, 2016).

There is a consensus in the field concerning the relative usefulness and value of integrating risk state and dynamic risk factors into the assessment process to enhance case management (Serin et al., 2016). The incorporation of dynamic risk factors to guide case formulation and treatment selection "is now standard clinical practice" (Heffernan & Ward, 2015, p. 250; see also, Ward & Beech, 2015). Although dynamic risk factors have received relatively less attention and examination in risk assessment research, professionals have posited that the accurate identification and consideration of dynamic risk factors is important for three predominant reasons: (a) the sound and comprehensive assessment of risk for future violence; (b) the identification of treatment targets; and (c) the monitoring of treatment progress, change in dynamic risk factors, and change in overall risk state (Cording et al., 2016; Klepfisz et al., 2016; McNaughton Nicholls & Webster, 2014).

Consideration of dynamic risk factors is essential to conduct a comprehensive evaluation of an individual's risk for violence, especially if the ultimate aim of the assessment is the mitigation or reduction of risk (Dexter & Vitacco, 2020; Douglas & Skeem, 2005; Guy et al., 2015; Mills, 2017; Monahan & Skeem, 2016; Ward, 2016). Risk assessments that attempt to explain or justify the foundation and level of risk are superior at meeting the needs of decision makers and protecting those being assessed (Mann et al., 2010). These assessments provide greater clarity and assistance to decision makers by resting professional opinions on solid understanding and reasoning. Although useful for the prediction of violence, static risk factors offer little in the form of understanding an individual's risk. The sole reliance on static risk factors places serious limitations on the integration of information into a cohesive account of the individual's risk (i.e., the type, severity, frequency, likelihood, rationale, and potential victims). Including dynamic risk factors in the assessment process offers more fruitful information upon which to create an understanding of the individual's risk and elucidate this understanding through case formulation (Hart & Logan, 2011).

Dynamic factors are extremely promising not merely for their ability to predict violence or general recidivism, but also for their usefulness in identifying effective

treatment targets and guiding the selection of appropriate risk management strategies more generally (Cording et al., 2016; Klepfisz et al., 2016; McNaughton Nicholls & Webster, 2014). An individual's risk status or overall risk level as determined exclusively based on static risk factors is "philosophically deterministic and of no practical benefit when attempting to direct or monitor progress" for risk reduction (Jones et al., 2010, p. 860; see also Brown et al., 2009; Douglas & Skeem, 2005; Skeem & Mulvey, 2002). In contrast, numerous professionals have emphasized how the identification of relevant dynamic risk factors provides a direct link to determining reasonable targets for intervention strategies aimed at reducing the likelihood of future violence or offending (Andrews & Bonta, 2010b; Klepfisz et al., 2016; McNaughton Nicholls & Webster, 2014). In fact, identifying dynamic risk factors is considered "essential for effective intervention" (Douglas & Skeem, 2005, p. 351; see also Kraemer et al., 1997).

A primary justification for the focus on dynamic risk factors is their ability to inform treatment targets, which is founded in an overarching theory of effective correctional intervention (Andrews & Bonta, 2010b; Klepfisz et al., 2016). The Risk-Need-Responsivity (RNR) model of effective correctional intervention was developed over two decades ago based on numerous meta-analytic investigations and has since experienced wide-spread acceptance in the forensic and correctional fields (e.g., Andrews & Bonta, 2010a, 2010b; Andrews et al., 1990; Andrews & Dowden, 2007; Dowden & Andrews, 2000). The RNR models describes three main principles for effective correctional programming (i.e., risk management strategies aimed at reducing offending). One of these principles, the need principle, states that effective interventions are those that target criminogenic needs or dynamic risk factors. Several meta-analyses have observed that treatments targeting criminogenic needs are more effective at reducing (violent) recidivism compared to treatments that target other factors (e.g., Andrews, 2012; Andrews & Bonta, 2010b; Dowden & Andrews, 1999a, 1999b, 2000, 2003, 2004; Hanson & Morton-Bourgon, 2009; Koehler et al., 2013). Criminogenic needs are described as the "targets of change" that management strategies need to address to reduce the chances of offending (Andrews & Bonta, 2010a, p. 45).

The consideration of risk state and dynamic risk factors is also vital for monitoring change and establishing whether meaningful progress has been made against the identified treatment targets (Cording et al., 2016; Douglas & Skeem, 2005; Klepfisz et al., 2016; McNaughton Nicholls & Webster, 2014). Professionals are tasked with not only

identifying individuals at greatest risk of violence and recommending strategies to reduce this risk, but also with determining to what extent an individual's level of risk (state) has changed. The inherent link between dynamic risk factors and monitoring change or progress in treatment is clear; "a crucial component of the concept of dynamic risk is that it is just that – dynamic, or in other words, changeable" (Cording et al., 2016, p. 93). Through the inclusion of dynamic risk factors coupled with repeated assessments, professionals can monitor an individual's risk state and any impacts on this state from intervention or management strategies. Moreover, it has been noted that the schedule of monitoring needs to approximate the rate of change in the given dynamic risk factors, otherwise the monitoring "will miss ebbs and flows in risk state – including those associated with intervention" (Douglas & Skeem, 2005, p. 351).

In addition to the above three reasons that consideration of dynamic factors is imperative (i.e., comprehensive assessment of risk, identification of treatment targets, and monitoring change and treatment progress), Viljoen, Gray and colleagues (2017) developed three hypotheses to explain the importance of reassessments and, in particular, why reassessments based largely on dynamic risk factors might improve risk predictions. These hypotheses include the shelf-life hypothesis, the dynamic change hypothesis, and the familiarity hypothesis. It should be noted that this set of hypotheses were not developed or intended to be mutually exclusive nor exhaustive of all possible reasons for the importance of reassessments and why they may improve the predictive validity of risk assessments.

As described by Viljoen, Gray and colleagues (2017), the *shelf-life hypothesis* states that opinions of risk may become outdated, elapsed, or invalid over time. As a result, risk assessments may be better suited for short-term rather compared to long-term predictions of violence. On the other hand, the *dynamic change hypothesis* states that opinions of risk may be more accurate in repeated assessments compared to single or baseline assessments because professionals are able to track within-individual fluctuations in risk state. By allowing professionals to monitor individual changes in risk state and dynamic risk factors, opinions of risk can be altered and compensate for changes in dynamic risk factors. As individuals may exhibit a decrease (or increase) in the presence or relevance of dynamic risk factors and a corresponding reduction (or elevation) in the likelihood of perpetrating violence, reassessments might more accurately capture the individual's risk. Thus, both the shelf-life and dynamic change

hypothesis emphasize intra-individual changes in risk state. Conversely, the *familiarity hypothesis* states that professionals may acquire additional information from each subsequent reassessment and this new information may lead to improvements in the ability to predict violence. The familiarity hypothesis, thus, emphasizes changes in the knowledge of the evaluator regarding the given individual.

Based on the above-described rationale, the importance of risk state and dynamic risk factors are apparent. As such, “the field’s next greatest challenge is to develop sound methods for assessing changeable aspects of violence risk and systematic methods for targeting these aspects to reduce violence” (Douglas & Skeem, 2005, p. 349; see also Dvoskin & Heilbrun, 2001). Although it is now standard practice to include purportedly dynamic risk factors in a comprehensive risk assessment, the use of these tools to assess change is relatively neglected. The development of sound tools to assess change in risk factors and risk state, as well as using this data to guide risk management strategies, “can not only improve lives and reduce sexual violence but is an ethical and human responsibility” (Olver & Stockdale, 2020, p. 2).

1.4. Measuring Change in Dynamic Risk

Accurately and reliably measuring change is critical, as it may assist professionals at improving both risk prediction and risk prevention (Viljoen, Gray, et al., 2017). In 2005, Douglas and Skeem noted that the field has “yet to identify which risk factors for violence are modifiable and how to assess them on an ongoing basis” (p. 367). Over the ensuing years, a number of purportedly dynamic risk factors have been identified and incorporated into decision support tools. Nevertheless, it is still the case that measuring change in violence risk assessments is relatively novel and “accomplishing this empirical task is necessary to inform violence management” (Douglas & Skeem, 2005, p. 367).

However, there is a relative dearth of empirical examinations regarding the ability of and extent to which contemporary risk assessment instruments are capable of measuring change in risk factors and risk state (Douglas & Skeem, 2005; Viljoen, Gray, et al., 2017). In order to provide targeted intervention strategies aimed at reducing violence, the criminal justice and mental health systems require risk assessment instruments that can reliably and validly identify treatment targets (i.e., criminogenic

needs), as well as measure changes in these needs that will impact each individual's likelihood of violence (Ainsworth & Taxman, 2013; Bonta, 2002; Caudy et al., 2013; Olver & Stockdale, 2020; Skeem & Monahan, 2011).

The measurement of dynamic risk factors and change over time in these factors is complex (Beech et al., 2016). So as to ensure that the most appropriate evaluations of a measure's ability to change over time are being taken into account, it is necessary to articulate several psychometric properties that are of particular relevance to evaluating change (Guyatt et al., 1987; Freeman et al., 2013; Husted et al., 2000; Viljoen, Shaffer, et al., 2017). Each of these properties will be reviewed in turn, highlighting their applicability to evaluating the main criteria for determining whether a risk factor is in fact a dynamic risk factor (i.e., a causal risk factor or a criminogenic need).

Viljoen, Shaffer and colleagues (2017) outlined a framework for gauging the ability of a risk assessment instrument to measure change over time. This framework was developed based on the combination of prior literature in the field of risk assessment (e.g., Douglas & Skeem, 2005; Monahan & Skeem, 2016) with more general literature regarding treatment outcomes, clinically significant change, and individual change statistics (e.g., Beaton et al., 2001; Duff, 2012; Freeman et al., 2013; Guyatt et al., 1987, 1989; Husted et al., 2000; Lambert & Vermeersch, 2013; McGlinchey et al., 2002; Riddle & Stratford, 2013; Speer & Greenbaum, 1995; Wright & Young, 1997). Under this framework, three psychometric properties are acutely relevant for appraising the extent to which a risk assessment tool can detect change: measurement error, responsiveness, and utility (see Table 1-2).

Measurement error is defined as the difference between a true score and a measured score or, relatedly, the proportion of a score that is due to error (Bland & Altman, 1996; Buonaccorsi, 2010; Duff, 2012; Viljoen, Shaffer, et al., 2017; Weems, 2007). In the context of examining change on a risk assessment tool, measurement error refers to the extent to which a tool can measure change using a reliable method or the extent to which change in scores over time are due to real changes in the underlying characteristic rather than simply due to measurement error (Viljoen, Shaffer, et al., 2017). In the context of a single assessment, error can reflect any systematic or random bias in the measurement, and over multiple assessments, these biases can be intensified or exacerbated (Duff, 2012).

Table 1-2. Framework for Evaluating Change in Dynamic Risk

Descriptor	Definition	Source
Measurement error	Difference between a true score and a measured score; Variability in measurements of the same quantity on the same subject; Proportion of a score that is attributable to error	Bland & Altman (1996) Buonaccorsi (2010) Duff (2012)
Responsiveness	Ability of an index to measure (clinical or significant) change when it has occurred	Beaton et al. (2001) Guyatt et al. (1987) Wright & Young (1997)
Internal responsiveness	Ability of an index to measure change over a specified time period	Husted et al. (2000)
External responsiveness	Extent to which change on an index relates to change on an external criterion or reference index	Husted et al. (2000)
Comparative responsiveness	Relative extent of an index to measure change compared to another index	Freeman et al. (2013)
Utility	Extent to which change on an index improves on typical decision-making and relevant outcomes	Hunsley & Bailey (1999) Hunsley (2003)

Note. Adapted from Viljoen, Shaffer, et al. (2017).

Across numerous contexts, a significant challenge when examining change is the ability to separate real change from measurement error, as any observed change in scores is likely an unknown combination of true change and error (Duff, 2012; Jacobson & Truax, 1991). For example, if an individual exhibits a small reduction on a purported dynamic risk factor, this reduction may be the result of actual change on the factor, or it may result from imperfect reliably in the measurement of the factor. Accordingly, the consideration of measurement error and overall reliability of measurements obtained using an instrument is a necessary step in examining the ability of a risk assessment tool to detect changes in risk and risk factors over time. In order for a risk assessment instrument to be able to detect reasonably small changes in risk, measurement error must be satisfactorily low, and the measurements should exhibit relatively rare ceiling and/or floor effects (Viljoen, Shaffer, et al., 2017).

A second critical psychometric property for examining the ability of a risk assessment instrument to detect change over time is *responsiveness* (Viljoen, Shaffer, et al., 2017). Generally, responsiveness is defined as the ability of an instrument to capture real change that occurred over time (Beaton et al., 2001; Freeman et al., 2013; Husted et al., 2000; Wright & Young, 1997). That is, responsiveness attempts to measure the relationship between change on a score and real change on the underlying characteristic (Freeman et al., 2013). As such, responsiveness is in part dependent upon the measurement error of the given index. In the context of examining change on a risk assessment tool, responsiveness refers to the extent to which a risk assessment tool can detect true changes that have occurred over time or if the tool is sensitive to change (Viljoen, Shaffer, et al., 2017). Notably, some authors have discussed responsiveness under the label sensitivity or sensitivity to change, with each of these terms being used synonymously. However, sensitivity to change (i.e., responsiveness) should not be confused with sensitivity in the context of the accuracy of classification systems or diagnostic tests (i.e., sensitivity – the true positive rate – versus specificity – the true negative rate; Viljoen, Shaffer, et al., 2017).

Several authors have distinguished between different features or types of responsiveness (see Table 1-2). According to Husted and colleagues (2000), there are two key forms of responsiveness: internal and external responsiveness. *Internal responsiveness* describes “the ability of a measure to change over a particular prespecified time frame” (p. 459). A common method for evaluating internal

responsiveness involves examining the extent to which a measure detects changes across different types or levels of treatment. A tool is considered to exhibit internal responsiveness if it detects greater change following an efficacious treatment or is able to differentiate between different levels of treatment based on the extent of change on the measure. As such, a risk assessment instrument should display decreases in risk factor scores and increases in protective factor scores after an effective treatment (Viljoen, Shaffer, et al., 2017). In addition to treatment-related changes, some measures may change spontaneously as a result of major life events or developmental processes. Thus, regardless of treatment involvement, a risk assessment instrument should detect some changes in risk factors and risk state over time. As well, internal responsiveness is dependent upon to the context, including the specific methodology, measure(s), treatment(s) or lack thereof, and outcome(s) used in the study (Husted et al., 2000).

External responsiveness describes "the extent to which changes in a measure over a specified time frame relate to corresponding changes in a reference measure" (Husted et al., 2000, p. 459). In contrast to internal responsiveness which focuses on the ability of an index or scores to change over time, external responsiveness focuses on the relationship between changes on this index (or scores) and change in some external criterion of interest. In this sense, internal responsiveness is concerned only with the measure itself, whereas external responsiveness is concerned with the relationship between the tool and an external standard. In the context of risk assessment instruments, external responsiveness would be evident if a measure exhibits lower rates of violence following a decrease in dynamic risk factors or higher rates of violence following an increase in dynamic risk factors (Viljoen, Shaffer, et al., 2017). Thus, for a tool to be considered to demonstrate external responsiveness, evidence must exhibit a relationship between change on the measure and change on the relevant criterion, standard, or indicator. Unlike internal responsiveness, external responsiveness is argued to only be dependent upon to the selection of the appropriate external criterion. Implicit in this understanding is the notion that "external responsiveness is a property of a measure and therefore it has meaning in a wider range of settings than the more context specific concept of internal responsiveness" (Husted et al., 2000, p. 460).

Both internal and external responsiveness are directly relevant to validating a risk factor as a dynamic risk factor in the context of violence risk assessments. By definition a dynamic risk factor must (a) precede violence, (b) increase the propensity for violence,

(c) be capable of change over time, and (d) be capable of predicting changes in the likelihood of violence when change occurs on the factor (Andrews & Bonta, 2010b; Douglas & Skeem, 2005; Klepfisz et al., 2016; Kraemer et al., 1997). Internal responsiveness speaks directly to the changeability of a factor over time; a purportedly dynamic risk factor that has been shown empirically to capture change over time would exhibit internal responsiveness. Thus, the factor would satisfy the third criteria of a dynamic risk factor. As well, risk state assessed using a decision support tool should also demonstrate internal responsiveness (i.e., empirical evidence of changeability over time). External responsiveness speaks directly to the extent to which changes on a factor are associated with changes in a criterion; a purportedly dynamic risk factor that has been shown empirically to change and this change is empirically associated with a change in the likelihood of violence would exhibit external responsiveness. Thus, the factor would satisfy the fourth criteria of a dynamic risk factor. In a similar manner, risk state should not only change over time, but should be associated with changes in the likelihood of violence (i.e., exhibit external responsiveness).

Additionally, *comparative responsiveness* describes the relative ability to detect change in one index compared with another index (Freeman et al., 2013). Certain features or characteristics are believed to be more changeable (e.g., greater frequency, speed, or magnitude of change) than others, and measures of these features should demonstrate greater comparative responsiveness. In other words, measures designed to assess dynamic factors should capture greater change than measures designed to assess fixed or stable constructs (Viljoen, Shaffer, et al., 2017). For example, some risk assessment tools are hypothesized to measure acute dynamic risk factors, whereas other assessment tools measure relatively stable traits. Thus, it is expected that more change would be seen on the measures aimed at purportedly more changeable characteristics, and there would be a difference in the measures' comparative responsiveness.

As such, comparative responsiveness is directly relevant to differentiating between static risk factors, stable dynamic risk factors, and acute dynamic risk factors. The main distinction between these types of risk factors rests on their relative speed and frequency of change over time; acute dynamic risk factors are capable of change quickly and over short period of time compared to the more gradually changing stable dynamic risk factors and the unchanging static risk factors. Thus, acute dynamic factors should

demonstrate the greatest comparative responsiveness compared to stable dynamic risk factors which in turn should demonstrate more comparative responsiveness than static risk factors. Moreover, comparative responsiveness allows for the evaluation of risk state assessed based on decision support tools that are designed to measure factors in briefer time frames compared to tools for assessing long-term risk.

Finally, *utility* refers to "whether the obtained assessment data improve on typical clinical decision making, treatment outcome, and/or the cost-benefits ratios associated with making accurate clinical determinations" (Hunsley, 2003, p. 443; see also Hunsley & Bailey, 1999). In the context of violence risk assessments, utility describes the extent to which assessing change on the measure aids in legal decision making (e.g., intervention planning, sentencing, release decisions) and protection of the public (e.g., reduced violence and other negative outcomes; Viljoen, Shaffer, et al., 2017). Utility is rather complex in comparison to responsiveness. In order to assess the utility of a specific measure, it requires the estimation and weighting of a number of elements including the costs of completing the measure, improvements in decisions made with the measure, changes in the rates of false positive and false negatives when the measure is included, as well as the economic, psychological, and legal costs associated with these errors (Hunsley, 2003). Based on consideration of the totality of these elements, it can be determined whether there is value in incorporating a measure into practice.

Each of these psychometric properties, including the particular types or aspects of responsiveness, has several potential options for methodological and statistical examination (Beaton et al., 1997, 2001; Freeman et al., 2013; Husted et al., 2000). Beaton and colleagues (1997) noted "there is no 'gold standard' for summarizing responsiveness" (p. 89). The general lack of agreement on the specific definition of responsiveness and optimal evaluation strategies has led to a "proliferation of responsiveness statistics, with investigators often reporting several within one study" (Husted et al., 2000, p. 460). As the most appropriate responsiveness statistic is not only debated but also dependent upon the specifics of the methodology and measurement, "there is no single agreed method for assessing responsiveness: a variety of statistical techniques are used that complement one another" (Freeman et al., 2013, p. 1634).

1.5. Empirical Research on Dynamic Risk

Although substantial theoretical advancements have been seen in the field over the past several decades, professionals have argued that there remains inadequate direction for clinicians responsible for the ongoing treatment and monitoring violence risk (Belfrage & Douglas, 2002; Douglas, Blanchard, et al., 2013; Douglas & Skeem, 2005). The theoretical developments surrounding dynamic risk and risk state must be paralleled with equally sound empirical evidence. Belfrage and Douglas (2002) previously noted that the field's understanding of risk as dynamic "has, to date, been more conceptual and theoretical than empirical" (p. 25). Over the ensuing years, a number of empirical investigations into the assessment of dynamic risk factors and the nature of change on purportedly dynamic risk factors have been reported (e.g., Andrews & Bonta, 2010b; Beggs & Grace, 2011; Coupland & Olver, 2018; Greiner et al., 2015; Hogan & Olver, 2016, 2019; Kroner & Yessine, 2013; Olver & Stockdale, 2020; Penney et al., 2016). Nevertheless, it is still the case that the "there is much more empirical direction for assessing risk than for reducing risk" (Monahan & Skeem, 2016, p. 508; see also Olver & Stockdale, 2020).

Other researchers have yet to explicitly use the framework developed by Viljoen, Shaffer, et al. (2017; i.e., internal, external, and comparative responsiveness) when investigating change on structured risk assessment tools. However, a growing body of empirical studies can speak to these crucial aspects of responsiveness. Alternatively, studies examining dynamic risk factors may be classified as using single time-point evaluations, dual time-point evaluations, or multiple time-point evaluations (Douglas & Skeem, 2005). However, these different methodologies are not equivalent in their soundness or utility for investigating change.

Single time-point evaluations of purportedly dynamic risk factors cannot investigate change and should not be regarded as proper examinations of dynamic risk, regardless of their intentions or claims (Beech et al., 2002; Dempster & Hart, 2002; Douglas & Skeem, 2005; Thornton, 2002). In order to truly investigate change, at least two measurements of risk are needed (Coupland & Olver, 2018; Serin et al., 2016). Although dual time-point evaluations can examine change over time, these studies are limited in their ability to examine different types and frequencies of change, in that only linear change can be inferred from two measurement points and speed of change is

limited to the timing of the measurements (Klepfisz et al., 2016). Thus, most scholars advocate that multiple time-point evaluations (i.e., three or more assessments) are best, as they allow for the examination of different trajectories and patterns of change, as well as separate true change from measurement error (Brown et al., 2009; Greiner et al., 2015; Hanson et al., 2007; Klepfisz et al., 2016; Quinsey, Jones, et al., 2006; Jones et al., 2010; Serin et al., 2016; Singer & Willett, 2003).

It must be noted that a grand body of literature exists concerning the examination of putatively dynamic risk factors operationalized in various manners and measured using various psychometric tools (e.g., Barnett et al., 2013; Bowen et al., 2008; Brown et al., 2009; Kingston & Olver, 2018; Jones et al., 2010; Lussier & Davies, 2011; Olver et al., 2014; Wakeling et al., 2013; Woessner & Schwedler, 2014). These studies have investigated the ability of various purported dynamic risk factors to change over time, and a smaller number of studies have also examined the association between change on these dynamic factors and change on external negative outcomes (e.g., violence and offending). For instance, in their influential article, Douglas and Skeem (2005) reviewed the available empirical support for several dynamic risk factors. Their review concluded that there is empirical support for the changeability of these factors and a smaller number of studies have found this change is associated with change in the likelihood of violence. Olver and Stockdale (2020) provide a comparable review of changeability in psychological characteristics in sexual offenders. Overall, their review summarizes the inconsistency across the empirical studies, with some studies finding evidence of changeability and a smaller number associating this change with change in recidivism. In a similar vein, Serin and colleagues (2013) conducted a comprehensive review of studies examining the relationship between intraindividual change and reductions in recidivism. Across the 49 effect sizes identified from relevant studies, they found 26 (53.1%) of these effects evidenced a relationship between change on a single dynamic risk factor and subsequent change in recidivism. The dynamic risk factors exhibiting a relationship between change and recidivism were almost exclusively from variables identified as the most important risk factors for general recidivism.

However, these studies have limited applicability and generalizability to the assessment of intraindividual change using structured risk assessment instruments. Internal responsiveness is contingent upon a number of methodological and statistical factors, including the measurement of the variables (Freeman et al., 2013; Husted et al.,

2000). Although these studies have found some evidence when measured as unitary constructs using various psychometric instruments, these results may not hold true for dynamic risk factors on established risk assessment protocols. There is considerable variation in the operationalization and measurement of dynamic risk factors between psychometric instruments and risk assessment instruments. As Serin and colleagues concluded from their systematic review of the literature, there is a need to "utilize better measurement strategies" (2013, p. 32). Accordingly, and consistent with testing standards (AERA, APA, NCME, 2014), it is necessary to empirically investigate the responsiveness of dynamic risk factors as they are used in clinical practice. That is, researchers should investigate the changeability of dynamic risk factors over time, and the association of this change with change in relevant outcomes, using risk assessment tools as they are meant to be used in practice.

As such, the current review of the empirical literature concentrates on the examination of putatively dynamic risk factors[1] and risk state using two prominent structured risk assessment instruments: the Historical-Clinical-Risk Management-20 (HCR-20; Webster et al., 1997) and the Short-Term Assessment of Risk and Treatability (START; Webster et al., 2009; see the Measures section for descriptions and Table A-1 for a list of items). The following sections review the empirical examinations relevant to each of the criteria for establishing a feature or characteristic as a dynamic risk factor: precedence and association, internal responsiveness, and external responsiveness, respectively. As well, the empirical investigations of changeability in risk state (i.e., internal and external responsiveness) using these tools are reviewed.

1.5.1. Precedence and Association

A vast body of empirical literature has investigated whether dynamic risk factors are in fact risk factors; that is, whether these characteristics precede violence and are associated with an increased risk of perpetrating violence. These lines of research have generally relied on single time-point research designs (i.e., assessing risk factors once and subsequently measuring violence; Douglas & Skeem, 2005). The field is currently

[1] From this point on, items included on structured risk assessment tools that are labeled as and theoretically deemed to be dynamic shall be referred to as *dynamic risk factors*. And scales on structured risk assessment tools consisting solely of such items, shall be referred to as *dynamic risk scales*, consistent with the labeling and intention of the risk assessment tools.

replete with independent empirical investigations demonstrating that dynamic risk factors on structured risk assessment protocols are associated with and predictive of violence (Andrews & Bonta, 2010b; Brown et al., 2009; Douglas et al., 2014; Gendreau et al., 1996; Douglas & Otto, 2020). Copious empirical support has been found for the predictive validity of both risk assessments tools comprised of a combination of static and dynamic risk factors (e.g., Caudy et al., 2013; Daffern & Howells, 2007; Dolan & Fullam, 2007; Douglas et al., 2003; Douglas, Cox, et al., 1999; Douglas et al., 2005), as well as risk assessment tools comprised solely of dynamic risk (and protective) factors (e.g., Chu et al., 2011; Daffern et al., 2009; Grann et al., 2005; Nonstad et al., 2010). Studies have found that dynamic risk factors are predictive of various types and measurements of violence in institutional (e.g., Belfrage et al., 2000; Chu et al., 2011; Daffern & Howells, 2007; Daffern et al., 2009; Dolan & Fullam, 2007; Vitacco et al., 2012) and community settings (e.g., Douglas et al., 2003; Douglas, Ogloff, et al., 1999; Douglas et al., 2005; Simourd, 2004; Wong & Gordon, 2006). Overall, this research supports the ability of dynamic risk factors to predict violence and recidivism "with moderate to high levels of accuracy" (Jones et al., 2010, p. 861).

Numerous studies have also demonstrated that dynamic risk factors add incrementally to the predictive validity of static risk factors (e.g., Beech et al., 2002; Beggs & Grace, 2011; Craissati & Beech, 2003; Dempster & Hart, 2002; Hanson & Harris, 2000). Again, this is based predominantly on single time-point evaluations. That is, dynamic risk factors have been shown to improve upon the statistical association between static risk factors and violence. In particular, studies have found that scales consisting of dynamic risk factors on structured risk assessment instruments add above and beyond the predictive validity of scales consisting of static risk factors on the same instruments (e.g., Beggs & Grace, 2011; Vincent et al., 2011). Relatedly, studies have found that tools and scales consisting solely of dynamic risk factors add above and beyond the statistical relationship between well-validated risk assessment protocols and violence (e.g., Desmarais et al., 2012; Doyle & Dolan, 2006; Eher et al., 2012). In general, the research supports the view that "dynamic risk factors predict recidivism as well as or better than static, historical variables" (Hanson & Harris, 2000, p. 7), but both static and dynamic risk factors are useful for predicting violence (Dexter & Vitacco, 2020; Vitacco et al., 2012).

Additionally, a number of meta-analyses have confirmed the results of these independent studies (e.g., Campbell et al., 2009; Eisenberg et al., 2019; Gendreau et al., 1996; Guy, 2008; van den Berg et al., 2018; Yang et al., 2010) and a number of qualitative reviews are available (e.g., Craig et al., 2005; Craissati & Beech, 2003; Douglas & Skeem, 2005; Olver & Stockdale, 2020; Otto, 2000; Serin et al., 2013). Meta-analyses have demonstrated that dynamic risk factors and dynamic scales perform at least as well as static risk factors and static scales in predicting violence and recidivism (Gendreau et al., 1996). For instance, based on a meta-analysis of 27 independent studies that investigated the predictors of general and violent recidivism, including 543 effect sizes over 116,982 forensic outpatients, Eisenberg and colleagues (2019) observed that all of the included risk factors predicted recidivism with small to moderate effect sizes. However, dynamic risk factors were more strongly predictive of general and violent recidivism than static risk factors. Similarly, based on a meta-analysis of 52 studies (N = 13,446) examining the predictive validity of static and dynamic risk instruments in adult male sex offenders, van den Berg and colleagues (2018) found that dynamic risk assessment protocols are predictive of sexual (d = 0.71) and violent (d = 0.43) recidivism, as well as add incremental predictive validity to static risk assessment protocols. Consistent with previous meta-analyses (e.g., Hanson & Morton-Bourgon, 2009), the largest predictive effects were found for the specific outcome that the risk assessment tool was designed to assess.

Relatedly, several meta-analyses have found relatively little to no difference between the predictive capabilities of well-validated risk assessment instruments that comprise static and/or dynamic risk factors (Campbell et al., 2009; Hanson & Morton-Bourgon, 2009; Yang et al., 2010; Singh & Fazel, 2010). For instance, Yang and colleagues (2010) examined and compared the predictive abilities of nine commonly used risk assessment tools and their subscales. Based on 28 independent reports, they concluded that apart from one scale (i.e., the Psychopathy Checklist Factor 1) all the tools predicted violence with little difference in the predictive efficacy of the various instruments. That is, most contemporary risk assessment protocols with a focus on dynamic risk factors appear to achieve relatively comparable predictive validity, “with moderate rates of predictive validity for these measures in predicting sexual, violent, and general recidivism” (Cording et al., 2016, p. 88; see also Campbell et al., 2009; Hanson & Morton-Bourgon, 2009). On the other hand, a systematic review and meta-analysis by

Singh et al. (2011) reported small differences in predictive validity dependent on the outcome of interest between nine of the most commonly used risk assessment protocols based on 68 studies including 25,980 participants.

Although the predictive validity of dynamic risk factors and measures has been clearly demonstrated in prior research, satisfying the two required criteria for labeling a characteristic as a risk factor, there remains questions and debate regarding the generalizability of these factors and tools to diverse populations (Cording et al., 2016). In particular, ambiguity remains surrounding how to appropriately operationalize these characteristics (Caudy et al., 2013; Taxman et al., 2013; Ward, 2015, 2016). Moreover, debate still exists regarding the extent to which these risk factors are salient across demographic variables such as age, gender, and ethnicity (Hannah-Moffat, 2004, 2009; Nicholls et al., 2015; Olver et al., 2014; Skeem, 2013; Van Voorhis & Presser, 2001; Van Voorhis et al., 2010).

Overall, the vast majority of studies examining the predictive and incremental validity of dynamic risk factors have used single time-point estimates of a theoretically dynamic risk factor (Douglas & Skeem, 2005; Kraemer et al., 1997; Olver & Stockdale, 2020; Serin et al., 2016). These studies can only speak to the first two criteria of a dynamic risk factors (i.e., a characteristic that precedes and is associated with an increase likelihood of the outcome) and, thus, cannot be considered true examinations of the changeability of these factors (Beech et al., 2002; Dempster & Hart, 2002; Douglas & Skeem, 2005; Thornton, 2002). In contrast, far fewer studies have concentrated on establishing that dynamic risk factors and dynamic risk scales exhibit the criteria that are essential to the field's conceptualization of dynamic risk: internal and external responsiveness (Klepfisz et al., 2016; Serin et al., 2016). As such, "it is not the dynamic risk factors per se that should be focused on in research, but rather the changes in these factors" (Eisenberg et al., 2019, p. 744).

1.5.2. Internal Responsiveness – Changeability

The third criterion for a dynamic risk factor is that the characteristic is capable of intraindividual change over time either spontaneously or as a result of intervention (Douglas & Skeem, 2005; Kraemer et al., 1997), also referred to as internal responsiveness (Husted et al., 2000). Overall, less attention has been given to this

criterion than the first two (Klepfisz et al., 2016). However, since Douglas and Skeem's (2005) recommendations more than 15 years ago, there has been increased attention to applied research on change properties using structured risk assessment instruments (Coupland & Olver, 2018).

As discussed above, the present review focused on the responsiveness of two structured risk assessment protocols, the HCR-20 and START. A larger body of empirical studies has evaluated the changeability of individual dynamic risk factors operationalized and measured in diverse manners. For instance, numerous longitudinal studies with children and adolescents have investigated trajectories of offending and some of these studies have also included examinations of changeability in risk factors that are associated with different trajectories (e.g., Brame et al., 2001; Kochenderfer-Ladd & Wardrop, 2001; Lussier & Davies, 2011; McDermott & Nagin, 2001; Nagin, 1999; Nagin & Tremblay, 1999; Osgood & Smith, 1995; Spieker et al., 1999). In addition, a number of studies have examined the extent to which individual dynamic factors measured using various instruments change over repeated measurements, as well as the degree to which this change is related to violence or recidivism (e.g., Barnett et al., 2013; Bowen et al., 2008; Brown et al., 2009; Hanson & Harris, 2000; Jones et al., 2010; Kingston & Olver, 2018; Lloyd et al., 2014; Odgers et al., 2009; Serin et al., 2009; Woessner & Schwedler, 2014). Studies have also examined the extent to which dynamic factors predict treatment progress and level of security (e.g., Hildebrand & de Ruiter, 2012; Müller-Isberner et al., 2007; Polaschek et al., 2016; Tengström et al., 2006).

Table 1-3 summarizes the empirical studies that have evaluated the internal responsiveness of the HCR-20 separated by type of research design. Overall, at least 19 studies have investigated the ability of dynamic risk factors and dynamic risk scales on the HCR-20 to change over time either spontaneously or as a result of intervention. Eight of the studies utilized a dual time-point design, whereas the other 11 employed a multiple time-point evaluation. Change on the HCR-20 has been investigated in samples of offenders, civil and forensic psychiatric patients, as well as combined samples. Notably, four of these studies used version three of the HCR-20 (Hogan & Olver, 2016, 2019; Mastromanno et al., 2018; Penney et al., 2016).

Table 1-3. Empirical Evaluations of Internal Responsiveness with the HCR-20

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Dual Time-Point Evaluations					
Belfrage et al. (2004)	47 (13)	Forensic psychiatric	Real-life assessments (in-vivo) by clinicians; Patients assessed at 3 to 24-month interval	Unreported	No change on C or R scales; No change on C or R items
Coupland & Olver (2018)	178	Violent offenders	Pseudo-prospective; Assessments by researchers based on files; Patients assessed at 6-month interval	Unreported; Percent change	Moderate to large change scores on C, R, and combined C-R scales; 14% change in one risk level on SPJ
de Vries Robbé et al. (2015)	108	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 5.6-year interval	Paired t-test	Mean change on H, C, & R scales; Mean change on total scores
Draycott et al. (2012)	29	Sexual/ violent offenders	Real-life assessments (in-vivo) by consensus by two multi-disciplinary professionals; Patients assessed at 2-year interval	Reliable change index	Reliable change in 31% of C scale; Reliable change in 20% of R scale
Hogan & Olver (2016)[a]	99	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval	Paired t-test	Mean change on C and R scales; Mean change on total scores
Hogan & Olver (2019)[a]	82	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval	Paired t-test	Mean change on C and R scales; Mean change on total scores
Mastromanno et al. (2018)[a]	40	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed at 4-month interval	Wilcoxon signed-rank test	Mean rank change on C and R scales; No change on combined C-R scale

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Sheldon & Gallagher (2010)	58	Forensic psychiatric	Pseudo-prospective; Real-life assessments (in-vivo) by clinicians; Patients assessed at unknown interval	Unreported	Mean change in C and R scales
Multiple Time-Point Evaluations					
Belfrage & Douglas (2002)	150 (70)	Forensic psychiatric	Real-life assessments (in-vivo) by clinicians; Patients assessed 3 times at 6-month intervals	Paired t-test; Wilcoxon signed-rank test	Mean change on C but not R scale; Mean change on 2 C items; Mean change on 2 R items
Blanchard (2013)	235	Offenders & civil psychiatric	Assessments by researchers using many sources; Patients assessed 6 times at 1-month intervals	Reliable change index	Reliable change in 5% of C scale; Reliable change in 6% of R scale
Douglas et al. (2011)	174	Forensic psychiatric	Assessments by unknown raters; Patients assessed 3 times at 6-month intervals	Repeated measures ANOVA	Mean change on C scale (R scale not reported)
Michel et al. (2013)	248 (153)	Civil & forensic psychiatric	Assessments by researchers using many sources; Patients assessed 5 times at 6-month intervals	Friedman's test; Percent change	Rank change on C and R scales; Rank change on 2 C items; Rank change on 4 R items; Change in 17-78% of ratings between any two assessments; Change in 42-100% of ratings between any five assessments
Morrissey et al. (2014)	532	Forensic psychiatric	Real-life assessments (in-vivo) by consensus by multi-disciplinary professionals; Patients assessed up to 6 times at 1-year intervals;	Mixed-effect modeling	Mean change on C and R scales; Mean change on total scores; Mean change of 1 on C and R scales

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Neves et al. (2010)	158	General offenders	Assessments by researchers using many sources; Patients assessed 3 times at 6-month intervals	Paired t-test; Percent change	Mean change on R but not C scale; No mean change on total scores; No mean change on any C items; Mean change on 1 R item; Change on 15% of SRJ
Nitschke et al. (2020a)	121	Forensic psychiatric	Quasi-experimental, two-factor mixed design; Assessments by consensus by two researchers; Patients assessed 3 times at 6-month intervals	Mixed-effect modeling	No mean change on C scale; Greater change in R scale in Tx group
Nitschke et al. (2020b)	122	Forensic psychiatric	Quasi-experimental, two-factor mixed design; Assessments by consensus by two researchers; Patients assessed 5 times at 6-month intervals	Mixed-effect modeling	No mean change on C scale; Greater change in R scale in Tx group
O'Shea & Dickens (2015)	480	Civil psychiatric	Pseudo-prospective; Real-life assessments (in-vivo) by clinicians; Patients assessed at least 2 and up to 4 times with an average 181-day intervals	Mixed-effect modeling; Reliable change index	Mean change on C but not R scale; Mean change on total scores; Mean change on 4 C items; No mean change on any R items; Reliable change in 6% of C and 5% of R scales across all four assessments; 1% decreased two and 10% decreased one risk level on SRJ
Olsson et al. (2013)	267	Forensic psychiatric	Pseudo-prospective; Real-life assessments (in-vivo) by consensus by multi-disciplinary professionals; Patients assessed at least 2 times with an average 9-month interval and a 43-month interval between the first and most recent assessments	Wilcoxon signed-rank test	Mean rank change on C and R scales; Mean rank change on 5 C items; Mean rank change on 2 to 5 R items

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Penney et al. (2016)[a]	87 (63)	Forensic psychiatric	Assessments by researchers using many sources; Patients assessed at baseline, 1-month, and 6-months post-discharge	Paired t-test	No mean change on C or R scales; Mean change on C & R scales in low risk group only (based on H scale); Increase in low risk on SRJ; Decrease in moderate risk on SRJ

Note. Adapted and updated from Viljoen, Gray, et al. (2017).
[a] These studies used the Historical-Clinical-Risk Management-20 Version 3.

Overall, the results of these 19 studies are promising. Nine studies reported generally consistent results that largely support the internal responsiveness of this measure (Coupland & Olver, 2018; de Vries Robbé et al., 2015; Douglas et al., 2011; Draycott et al., 2012; Hogan & Olver, 2016, 2019; Morrissey et al., 2014; Olsson et al., 2013; Sheldon & Gallagher, 2010). Another nine studies reported somewhat inconsistent results that only partially support the internal responsiveness of the HCR-20 (Belfrage & Douglas, 2002; Blanchard, 2013; Michel et al., 2013; Mastromanno et al., 2018; Neves et al., 2010; Nitschke et al., 2020a, 2020b; O'Shea & Dickens, 2015; Penney et al., 2016). Studies were concluded to have found partial support if they reported change in only certain analyses (i.e., on only one of the dynamic risk scales or a portion of the dynamic risk factors). Finally, only one study did not find any support for the changeability on this tool (Belfrage et al., 2004).

Only a small subset of these studies in fact examined change on dynamic risk factors, whereas most examined the changeability of dynamic risk scales (i.e., scales comprising a set of dynamic risk factors). Specifically, six studies examined change on the dynamic risk factors on the HCR-20 (all using version two). Olsson and colleagues (2013) reported the greatest extent of change. They observed change on all five items on the Clinical scale from the first to the most recent assessment, as well as three of the Risk Management items when rated for an institutional setting (i.e., IN ratings) and all five Risk Management items when rated for a community setting (i.e., OUT ratings). Alternatively, O'Shea and Dickens (2015) reported change on four of the Clinical items (all except Unresponsive to Treatment) and none of the Risk Management items across repeated assessments. Michel et al. (2013) found that change was dependent upon the sample type. In both samples, change was seen on two of the Clinical items and four of the Risk Management items, but the specific items that evidenced change were different across the samples. In the forensic psychiatric sample, change was seen on two Clinical items, Lack of Insight and Impulsivity, and all Risk Management items except Exposure to Destabilizers. In the civil psychiatric sample, change was seen on two Clinical items, Impulsivity and Unresponsive to Treatment, and all Risk Management items except Lack of Personal Support. Belfrage and Douglas (2002) reported change on some of the dynamic items: two Clinical items and two Risk Management items. Other studies found even less change. Neves and colleagues (2010) reported change on only one Risk

Management item and no Clinical items. Finally, Belfrage et al. (2004) reported no change in any items across assessments.

In contrast, most studies examined change on dynamic risk scales. A single study observed no change on either the Clinical or Risk Management scales (Belfrage et al., 2004). Another study found no change on either scale scores when examining the entire sample but reported mean level change on both scale scores in the low-risk group when the sample was split based on static risk factors (i.e., median split on the Historical scale; Penney et al., 2016). Others observed change on only one of the dynamic scales. Of the studies that examined both the dynamic risk scales on the HCR-20, some found change on only the Clinical scale but not the Risk Management scale (Belfrage & Douglas, 2002; O'Shea & Dickens, 2015), whereas others found change on only the Risk Management scale but not the Clinical scale (Neves et al., 2010; Nitschke et al., 2020a, 2020b). Comparatively, most studies found support for the internal responsiveness of both the Clinical and Risk Management scales (Coupland & Olver, 2018; de Vries Robbé et al., 2015; Hogan & Olver, 2016, 2019; Mastromanno et al., 2018; Michel et al., 2013; Morrissey et al., 2014; Olsson et al., 2013; Sheldon & Gallagher, 2010).

In addition, only a subset of these studies directly examined intraindividual change, whereas the remainder utilized analytical techniques that examined aggregate or group-level change. That is, only three of these studies examined change using a statistical analysis that can speak directly to intraindividual change (Blanchard, 2013; Draycott et al., 2012; O'Shea & Dickens, 2015). Amongst this subset of studies, varied results were evident. Draycott et al. (2012) reported the greatest extent of intraindividual change with reliable change seen in 31% of Clinical scale and 20% of Risk Management scale scores, compared to only 5% of Clinical scale and 6% of Risk Management scale scores reported by Blanchard (2013) and 6% of Clinical scale and 5% of Risk Management scale scores reported by O'Shea and Dickens (2015). One additional study reported the precent of scores that changed over assessment intervals. Michel and colleagues (2013) observed change in 17% to 78% of participants' dynamics items across any two assessments, as well as change in 42% to 83% of participants' dynamic items across all five assessments.

In contrast, most studies employed aggregate statistical techniques in the form of parametric and non-parametric tests on the mean or mean rank. These studies generally found full (de Vries Robbé et al., 2015; Hogan & Olver, 2016, 2019; Mastromanno et al., 2018; Michel et al., 2013; Morrissey et al., 2014; Olsson et al., 2013; Sheldon & Gallagher, 2010) or partial (Belfrage & Douglas, 2002; Neves et al., 2010; Nitschke et al., 2020a, 2020b; O'Shea & Dickens, 2015) support for the internal responsiveness of dynamic risk scales on the HCR-20. Although informative and group-level change implies individual-level change, these analytical techniques may obscure individual change by aggregating the data.

Very few of these studies examined change in overall ratings of risk state and can speak to the internal responsiveness of these ratings made using the HCR-20. That is, only four studies included analyses relevant to evaluating change in risk state (i.e., changes in overall ratings of risk or summary risk judgments). In general, all these studies found some support for the changeability of overall risk ratings. For instance, Neves and colleagues (2010) found the largest extent of change, observing change on 15% of the summary risk judgements. Coupland and Olver (2018) observed change of one risk level in 14% of ratings. O'Shea and Dickens (2015) reported slightly less change; they observed that 1% of participants decreased two levels (i.e., high to low risk) and 10% decreased one level (i.e., high to moderate risk or moderate to low risk). Finally, Penney et al. (2016) found an aggregate increase in the proportion of low-risk ratings (18% to 38%) and a comparable decrease in the number of moderate-risk ratings (76% to 54%) on the summary risk judgments.

Table 1-4 summarizes the empirical studies that have evaluated the internal responsiveness of the START separated by type of research design. Overall, at least six studies have investigated the ability of dynamic risk factors and dynamic risk scales on the START to change over time either spontaneously or as a result of intervention. Four of the studies utilized a dual time-point design, while the other two employed a multiple time-point evaluation. Change on the START has been investigated in samples of adolescent offenders (Sellers et al., 2017; Viljoen et al., 2012), forensic psychiatric patients (Hogan & Olver, 2016, 2019; Whittington et al., 2014), as well as a combined sample (Blanchard, 2013).

Table 1-4. Empirical Evaluations of Internal Responsiveness with the START

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Dual Time-Point Evaluations					
Hogan & Olver (2016)	99	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval	Paired t-test	Mean change on Strength and Vulnerability scales
Hogan & Olver (2019)	82	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval	Paired t-test	Mean change on Strength and Vulnerability scales
Sellers et al. (2017)[a]	291 (59)	Adolescent offenders	Real-life assessments (in-vivo) by clinicians; Patients assessed an average 115-day interval	Paired t-test; Reliable change index	No mean change on Strength scale; Mean change on Vulnerability scale; Mean change on 6 Strength items; Mean change on 4 Vulnerability items; Reliable change on 11% of Strength scale and 17% of Vulnerability scale
Viljoen et al. (2012)[a]	90 (81)	Adolescent offenders	Assessments by researchers using many sources; Patients assessed at 3-month interval	Paired t-test; Reliable change index	No mean change on Strength scale; Mean change on Vulnerability scale; Reliable change on 11% of Strength scale and 6% of Vulnerability scale; Change on 40% of SRJ

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Multiple Time-Point Evaluations					
Blanchard (2013)	235	Offenders & civil psychiatric	Assessments by researchers using many sources; Patients assessed 6 times at 1-month intervals	Reliable change index	Reliable change on 7% of Strength and 6% of Vulnerability scale
Whittington et al. (2014)	50	Forensic psychiatric	Pseudo-prospective; Real-life assessments (in-vivo) by nurses; Patients assessed a median number of 8 times (range = 1 to 25) at median 22-day intervals	Unreported	15% reduction in mean Strength and 10% reduction in mean Vulnerability scales across 17 assessments

Note: Adapted and updated from Viljoen, Gray, et al. (2017).
[a] These studies used the Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV; Viljoen et al., 2014).

Largely, the results of these six studies are promising in their support for the internal responsiveness of this measure, yet in need of further investigation. Three studies reported generally consistent results that support the internal responsiveness of this measure (Hogan & Olver, 2016, 2019; Whittington et al., 2014). Another three studies reported somewhat inconsistent results that only partially support the internal responsiveness of the START (Blanchard, 2013; Sellers et al., 2017; Viljoen et al., 2012). Studies were concluded to have found partial support if they reported change in only certain analyses (i.e., on only one of the dynamic scales or a portion of the dynamic risk factors).

Similar to the research on the HCR-20, most of the studies looking at the START examined changeability of dynamic scales, while only a single study examined change on dynamic risk and protective factors. Only Sellers and colleagues (2017) reported results for all the individual dynamic items included on the START. They observed change on a minority of the items: four of the Vulnerability items and six of the Strength items. Notably, this study investigated change on the adolescent version of the START (Viljoen et al., 2014). In contrast, the majority of studies examined change on dynamic risk scales. These studies are roughly equally split between those that found support for the internal responsiveness of both dynamic scales on the START (Hogan & Olver, 2016, 2019), and those that found support for the internal responsiveness of only one scale (Sellers et al., 2017; Viljoen et al., 2012). The remaining studies found some or limited evidence supporting the changeability of these scales (Blanchard, 2013; Whittington et al., 2014).

Once again, only a subset of the studies directly examined intraindividual change, whereas the remainder used analytical techniques that operate at the aggregate level. That is, only three of these studies examined change using a statistical analysis that can speak directly to intraindividual change (Blanchard, 2013; Sellers et al., 2017; Viljoen et al., 2012). Amongst this subset of studies, the results were generally modest. The two studies evidencing the greatest intraindividual change used the adolescent version of the tool. Sellers and colleagues (2017) reported the greatest extent of intraindividual change with reliable change seen in 17% of the Vulnerability scale and 11% of the Strength scale scores, while Viljoen and colleagues (2012) found slightly lower rates of reliable change (6% of the Vulnerability scale and 11% of the Strength scale scores). The only study of this sort using the START found lower rates of change.

Blanchard (2013) found reliable change in 6% of Vulnerability scale and 7% of Strength scale scores. One additional study reported the percent of scores that changed over assessment intervals. Whittington and colleagues (2014) observed a 10% reduction in mean vulnerability scale scores and a 15% reduction in mean strength scale scores across a median number of 8 assessments (up to a total of 17 assessments).

In contrast, the remaining studies employed aggregate statistical techniques in the form of parametric tests based on the mean (e.g., paired samples t-tests or repeated measures ANOVA). Although these studies did not directly examine intraindividual change, they generally found support for the internal responsiveness of both the vulnerability and strength scales (Hogan & Olver, 2016, 2019) or support for the internal responsiveness of the Vulnerability scale but not the Strength scale (Sellers et al., 2017; Viljoen et al., 2012).

Only one of these studies examined change in overall ratings of risk state and can speak to the internal responsiveness of these ratings made using the START. That is, only one study included analyses relevant to evaluating change in risk state (i.e., changes in the summary risk judgments). Overall, this study found support for the changeability of overall risk ratings. Using the adolescent version of the instrument with adolescent offenders, Viljoen and colleagues (2012) observed change on 40% of the summary risk judgements. No studies have investigated changeability in risk state using the original version of the START or with adult samples.

1.5.3. External Responsiveness – Linking Change

The final criterion for a dynamic risk factor is that change on the risk factor is associated with change in the likelihood of the outcome (Andrews & Bonta, 2010b; Douglas & Skeem, 2005; Kraemer et al., 1997), also referred to as external responsiveness (Husted et al., 2000). Many correctional and other management programs rest on the supposition that changes in dynamic risk factors will directly reduce an individual's propensity for violence (Beggs & Grace, 2011; Klepfisz et al., 2016). However, the least amount of empirical research has investigated this criterion using structured risk assessment protocols. Nevertheless, there has been increased attention to the external responsiveness of risk assessment measures (Coupland & Olver, 2018).

Table 1-5 summarizes the research that has examined the external responsiveness of the HCR-20 separated by type of research design. A total of at least 12 studies have investigated the extent to which change over time on the dynamic risk factors and dynamic risk scales included on the HCR-20 are associated with the likelihood of perpetrating violence. Five of the studies employed a dual time-point research design, while the remaining seven utilized a multiple time-point design. The external responsiveness of the HCR-20 has been investigated in samples of offenders (Coupland & Olver, 2018; Neves et al., 2010) and forensic psychiatric patients (Coid et al., 2015; de Vries Robbé et al., 2015; Douglas et al., 2011; Hogan & Olver, 2016, 2019; Mastromanno et al., 2018; Morrissey et al., 2014; Penney et al., 2016; Wilson et al., 2013), as well as combined samples (e.g., Blanchard, 2013; Michel et al., 2013). Notably, five of these studies used version three of the HCR-20 (Coid et al., 2015; Hogan & Olver, 2016, 2019; Mastromanno et al., 2018; Penney et al., 2016), while the remaining studies used version two.

In general, these 12 studies provide some support for the external responsiveness of this measure. However, the results are also quite varied (both across analyses within the same study and across different studies) and lacking in their ability to provide a clear picture of the relationship between changes on the tool and change in the perpetration of violence. Nevertheless, all but one of these studies found support for the external responsiveness of at least one rating on the HCR-20.

Roughly four of these studies reported results that support the external responsiveness of the HCR-20 (Blanchard, 2013; Douglas et al., 2011; Penney et al., 2016; Wilson et al., 2013). Another seven studies reported inconsistent results that provide partial support for the external responsiveness of the measure (Coid et al., 2015; Coupland & Olver, 2018; de Vries Robbé et al., 2015; Hogan & Olver, 2016, 2019; Michel et al., 2013; Neves et al., 2010). These studies tended to only find a relationship between change on certain dynamic risk factors or scales and the propensity for violence, or they reported an association between change on the tool and change in violence in some analyses. Finally, a single study did not find a relationship between changes on the HCR-20 and subsequent changes in violence (Mastromanno et al., 2018). Notably, this study contained several methodological limitations, including a small sample size, pseudo-prospective assessments based on hospital records from the first and last two months of admission, and violence based on official police reports.

Table 1-5. Empirical Evaluations of External Responsiveness with the HCR-20

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Dual Time-Point Evaluations					
Coupland & Olver (2018)	178	Violent offenders	Pseudo-prospective; Assessments by researchers based on files; Patients assessed at 6-month interval; Followed an average 9.7 years post-release	Cox regression survival analysis	Change on combined C-R scale predictive of violence and recidivism after controlling for total score; Change in SRJ predictive of general recidivism but not violent recidivism
de Vries Robbé et al. (2015)	108	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 5.6-year interval; Followed an average 11 years post-release	Logistic regression; AUC	Change on C and R scales not predictive of short-term violence; Change on C and R scales predictive of long-term violence after controlling for H scale and age
Hogan & Olver (2016)[a]	99	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval; Inpatient aggression recorded up to discharge	Logistic regression	Change on C, R or combined C-R scales not predictive of inpatient violence after controlling for length of stay and pretreatment total score; Change on C scale predictive of inpatient violence after controlling for length of stay and pretreatment C score; Change in combined C-R scale predictive of inpatient violence after controlling for length of stay and pretreatment C-R score

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Hogan & Olver (2019)[a]	82	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval; Followed an average 8.2 years post-discharge	Cox regression survival analysis	Change on combined C-R presence scale not predictive of either violence or recidivism after controlling for pretreatment total score; Change on combined C-R relevance scale predictive of violence but not general recidivism after controlling for pretreatment total score
Mastromanno et al. (2018)[a]	40	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed at 4-month interval; Followed a maximum of 12.8 years post-discharge	Logistic regression	Change on C and R scale not predictive of violence or recidivism, individually or after controlling for baseline risk
Multiple Time-Point Evaluations					
Blanchard (2013)	235	Offenders & civil psychiatric	Assessments by researchers using many sources; Patients assessed 6 times at 1-month intervals	Generalized estimating equations	Change on C and R scales predictive of violence
Coid et al. (2015)[a]	409	Forensic psychiatric	Assessments by researchers based on files and staff interviews; Patients assessed 3 times at 6-month intervals	Mixed level modeling	Repeated assessments on C and R scales predictive of violence; Repeated assessments on 4 C items and 4 R items predictive of violence
Douglas et al. (2011)	174	Forensic psychiatric	Assessments by unknown raters; Patients assessed 3 times at 6-month intervals	Repeated measures ANOVA	Change on C scale predictive of violence (R scale not reported)

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Michel et al. (2013)	248 (153)	Civil & forensic psychiatric	Assessments by researchers using many sources; Patients assessed 5 times at 6-month intervals	Generalized estimating equations	Change on R scale predictive of violence, but not C scale; Change on 3 C items and 3 R items predictive of violence
Neves et al. (2010)	158	General offenders	Assessments by researchers using many sources; Patients assessed 3 times at 6-month intervals	Paired t-tests	Recidivists' scores increased across assessments, while non-recidivists' scores showed no change
Penney et al. (2016)[a]	87 (63)	Forensic psychiatric	Assessments by researchers using many sources; Patients assessed at baseline, 1-month, and 6-months post-discharge; Followed an average of 34 months post-discharge	Cox regression survival analysis	Time-dependent scores on C and R scales predictive of violence after controlling for H scale
Wilson et al. (2013)	30	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed 4 times at 3-month intervals	Survival analysis	Time-dependent scores on combined C-R scale predictive of violence after controlling for H scale

Note. Adapted and updated from Viljoen, Gray, et al. (2017).
[a] These studies used the Historical-Clinical-Risk Management-20 Version 3.

Although these studies provide some support for the external responsiveness of the HCR-20, most of the studies examined the relationship between change on dynamic risk scales and change in violence. Only two studies examined the changeability of dynamic risk factors on this tool and the relationship to change in the propensity for violence. Michel and colleagues (2013) investigated the external responsiveness of the HCR-20 in a combined sample of civil and forensic psychiatric patients assessed five times at roughly six-month intervals. They found that change on three Clinical items (Negative Attitudes, Impulsivity, and Unresponsive to Treatment) and three Risk Management items (Plans Lack Feasibility, Lack of Personal Support, and Noncompliance with Remediation Attempts) were predictive of violence. Although three items from each of these dynamic risk scales were found to exhibit external responsiveness, only changes on the total Risk Management scale, not the Clinical scale, were associated with change in violence. Additionally, Coid et al. (2015) reported that repeated assessments on four Clinical and four Risk Management items were predictive of violence using version three of the HCR-20.

Most of these studies examined changes on dynamic risk scales. As mentioned, one of these studies found no association between changes on the Clinical scale or Risk Management scale and subsequent violence whether examined in isolation or after controlling for baseline level of risk (Mastromanno et al., 2018). On the other hand, a small number of studies found consistent support for the external responsiveness of both the Clinical and Risk Management scales (Blanchard, 2013; Penney et al., 2016). Other studies have reported inconsistent results depending on the analysis of focus. For instance, de Vries Robbé and colleagues (2015) found that changes on neither of the Clinical nor Risk Management scales were predictive of short-term violence, whereas changes on both Clinical and Risk Management scales were related to changes in long-term violence after controlling for static risk factors. Similarly, Hogan and Olver (2016) found that changes on Clinical or Risk Management scales, or the combination of the two dynamic scales, were not predictive of inpatient violence after controlling for length of inpatient stay and baseline HCR-20 total score. However, change on the Clinical scale was predictive of violence after controlling for length of stay and baseline Clinical scale score, but change on the Risk Management scale was not predictive of violence after controlling for length of stay and baseline Risk Management scale score. Finally, change

on the combined clinical and risk management scales was associated with changes in violence after controlling for length of stay and the combined baseline scores.

Additionally, some studies only examined the relationship between changes in the total dynamic risk items on the HCR-20 (i.e., combining the Clinical and Risk Management scale scores) and the perpetration of violence. For instance, Wilson and colleagues (2013) observed that time-dependent scores on the combined dynamic scales were associated with violence after controlling for baseline historical scale scores. Some of these studies also reported differences depending on the specific analysis and outcome being investigated. Coupland and Olver (2018) examined the relationship between changes on the combined Clinical-Risk Management scales and various indices of recidivism. With regards to community outcomes, after controlling for baseline HCR-20 total score, change on the combined dynamic scales was predictive of general recidivism, but was not predictive of violent recidivism. With regards to institutional outcomes, after controlling for baseline HCR-20 total score, change on the combined dynamic scales was associated with change in serious institutional recidivism, but was not predictive of any institutional recidivism. As well, Hogan and Olver (2019) found that change on the combined dynamic scales was not associated with any outcomes after controlling for baseline HCR-20 total score. However, they reported that change on the relevance ratings (made only on version 3 of the HCR-20) was predictive of violence but not recidivism after controlling for baseline HCR-20 (relevance) total scores.

These studies also used a diverse array of analytical techniques to investigate external responsiveness. Two studies examined the data using generalized estimating equations which is ideally suited for analyzing the relationship between change over time and an external criterion (Blanchard, 2013; Michel et al., 2013). Blanchard (2013) observed that change on both the Clinical and Risk Management scales was associated with violence, whereas Michel et al. (2013) found that only change on the Risk Management scale was predictive of violence. Other studies examined the relationship between change scores and violence using survival analyses. These studies similarly found varied results that partially support the external responsiveness of the HCR-20 (Coupland & Olver, 2018; Hogan & Olver, 2019; Penney et al., 2016; Wilson et al., 2013). Finally, a minority of studies used statistical procedures that relied on aggregate group-based comparisons (Douglas et al., 2011; Neves et al., 2010) and provided some support for the external responsiveness of this tool.

Notably, only a single study examined the external responsiveness of overall ratings of risk state made using the HCR-20. The vast majority of these studies focused on the numerical scores of the two dynamic scales included on this measure and failed to include any examination of the summary risk judgments. Using a pseudo-prospective design based on file reviews, Coupland and Olver (2018) examined the relationship between change in overall risk ratings and change in violence. The relationship was found to be dependent upon the outcome of interest. That is, change in the summary risk ratings was not predictive of violence, but was predictive of general recidivism.

Table 1-6 summarizes the research that has examined the external responsiveness of the START separated by research design. At least five studies have examined the extent to which change over time on the dynamic risk factors and dynamic risk scales included on the START are associated with the perpetration of violence. Two of the studies employed a dual time-point research design, while the remaining three utilized a multiple time-point design. The external responsiveness of the START has only been investigated in samples of forensic psychiatric patients (Hogan & Olver, 2016, 2019; Whittington et al., 2014; Wilson et al., 2013), as well as a combined sample (Blanchard, 2013).

Generally, these five studies provide some support for the external responsiveness of this tool; moreover, this support is accompanied by inconsistencies in the available results and a lack of data regarding certain features of the tool. Only one of these studies reported consistent support for the external responsiveness of the START (Whittington et al., 2014). The remaining four studies reported inconsistent results that partially support the external responsiveness of this measure (Blanchard, 2013; Hogan & Olver, 2016, 2019; Wilson et al., 2013). These studies, similar to the HCR-20, tended to only find a relationship between change on one of the two dynamic scales on the START and violence, only find a relationship with some outcomes but not others, or simply failed to include analyses of both the dynamic scales on the instrument.

Table 1-6. Empirical Evaluations of External Responsiveness with the START

Authors	Sample		Methodology		Results
	Size	Type	Procedure	Analyses	
Dual Time-Point Evaluations					
Hogan & Olver (2016)	99	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval; Inpatient aggression recorded up to discharge	Logistic regression	Change on Vulnerability scale predictive of violence after controlling for pretreatment Vulnerability scale; Strength scale not reported
Hogan & Olver (2019)	82	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed an average 19-month interval; Followed an average 8.2 years post-discharge	Cox regression survival analysis	Change on Vulnerability scale predictive of recidivism (but not violence) after controlling for pretreatment Vulnerability scale; Strength scale not reported
Multiple Time-Point Evaluations					
Blanchard (2013)	235	Offenders & civil psychiatric	Assessments by researchers based on many sources; Patients assessed 6 times at 1-month intervals	Generalized estimating equations	Change on Strength scale predictive of serious but not general violence; Change on Vulnerability scale not predictive of either type of violence
Whittington et al. (2014)	50	Forensic psychiatric	Pseudo-prospective; Real-life assessments (in-vivo) by nurses; Patients assessed a median number of 8 times (range = 1 to 25) at a median 22-day intervals	Logistic regression	10-point change on Vulnerability scale predictive of violence; 10-point change on Strength scale not predictive of violence
Wilson et al. (2013)	30	Forensic psychiatric	Pseudo-prospective; Assessments by researchers based on files; Patients assessed 4 times at 3-month intervals	Survival analysis	Time-dependent scores on Vulnerability but not Strength scale predictive of violence after controlling for H scale

Note. Adapted and updated from Viljoen, Gray, et al. (2017).

Although these studies provide some support for the external responsiveness of the START, none of these studies examined the relationship between change on individual dynamic risk (and protective) factors and violence. That is, all five studies focused on the scale level as opposed to individual items on the START; thus, there is currently no empirical evidence that can speak directly to the external responsiveness of specific dynamic factors on this instrument.

The available empirical studies can speak to the external responsiveness of the dynamic scales that comprise the START. All these studies found some evidence that supports the external responsiveness of this instrument, yet the totality of these studies presents a variable picture with many unanswered questions. Only one study reported that change on both the vulnerability and strength scales was predictive of subsequent violence, with a 10-point change on each scale being associated with changes in violence (Whittington et al., 2014). Other studies have reported inconsistent results depending on the analysis of focus. For instance, Hogan and Olver (2016) identified a relationship between change on the Vulnerability scale and change in violence, after controlling for baseline Vulnerability scale scores. In contrast, Hogan and Olver (2019) found that change on the Vulnerability scale was associated with change in general recidivism, but not violence, after controlling for baseline Vulnerability scale scores. Neither of these studies reported any data regarding the Strength scale. Wilson and colleagues (2013) found that time-dependent scores on the Vulnerability scale were predictive of violence after controlling for baseline HCR-20 historical scale scores; however, no association was found for the Strength scale scores. On the other hand, Blanchard (2013) reported near opposite results; change on the Strength scale was associated with serious violence but not general violence. However, change in Vulnerability scale scores was not associated with either type of violence.

Remarkably, none of the studies examining the external responsiveness of the START included investigations of overall risk state. All these studies focused on the numerical total scores of the two dynamic scales included on this measure and failed to include any examination of the summary risk judgments. As such, no empirical evidence can speak to the external responsiveness of global ratings of risk made using this tool.

1.6. Dynamic Risk and Dynamic Risk Factors

There is a dire need for clarity in the conceptualization and measurement of dynamic risk (Serin et al., 2016). Based on this review of the empirical research, it is evident that there is a relative sparsity of empirical work that has directly investigated the core features of dynamic risk. The majority of empirical investigations of dynamic risk are only relevant to explaining interindividual differences in risk for violence. There is a large body of empirical literature that supports the overall ability of purportedly dynamic risk factors and dynamic risk scales to not only predict violence in isolation but also add incrementally to the association between static risk factors and violence.

After more than 15 years since Douglas and Skeem's (2005) pivotal article, it remains true that most studies in this area "use single time-point estimates of a putatively dynamic construct" (p. 356; Coupland & Olver, 2018; Lussier & Davies, 2011; Olver & Stockdale, 2020; Olver & Wong, 2011). Although there is still some debate regarding the operationalization, measurement, and utility of these factors (e.g., Baird, 2009; Cording et al., 2016; Heffernan & Ward, 2015; Ward, 2015, 2016), the available research provides a strong argument for the inclusion of dynamic risk factors in comprehensive risk assessment protocols. However, this empirical evidence has little relevance to the distinguishing features of dynamic risk factors and risks state (i.e., internal and external responsiveness).

Although there is a relative lack of empirical research and inconsistency in the available findings, the available research does provide some support for the responsiveness of the HCR-20 and START. Responsiveness is a crucial feature of many measures that must be examined in detail. However, responsiveness is not considered a fixed property of a given instrument; it is dependent upon the population, context, and methodology (Beaton et al., 1997, 2001; Freeman et al., 2013; Husted et al., 2000). Therefore, it is unlikely for there to be a single estimate of the responsiveness of a given measure across all applications and populations. It is vital to accumulate evidence about the responsiveness of measures across different settings, contexts, and samples to assist professionals in selecting the most appropriate instrument for a given task (Freeman et al., 2013).

Overall, there is limited empirical research to date that has investigated the ability of dynamic risk factors on structured risk assessment instruments to change over time (i.e., internal responsiveness) and even less empirical research has investigated the extent to which this change is associated with change in the propensity for violence (i.e., external responsiveness; Douglas & Skeem, 2005; Hanson & Harris, 2000; Monahan & Skeem, 2014; Olver & Stockdale, 2020; Serin et al., 2013). Although a number of studies have found that dynamic risk factors and dynamic risk scales on these instruments are capable of measuring change (e.g., Belfrage & Douglas, 2002; Blanchard, 2013; Coupland & Olver, 2018; de Vries Robbé et al., 2015; Draycott et al., 2012; Hogan & Olver, 2016, 2019; Michel et al., 2013; Neves et al., 2010; O'Shea & Dickens, 2015; Sheldon & Gallagher, 2010; Whittington et al., 2014), relevant studies have also presented a murky picture of findings. It remains true that "the scientific focus on dynamic risk and risk management has been more conceptual than empirical" (Douglas & Skeem, 2005, p. 352).

From the available evidence, it is unviable to arrive at any definitive conclusions regarding the defining criteria for dynamic risk factors (Coupland & Olver, 2018; Klepfisz et al., 2016; Monahan & Skeem, 2014, 2016; Olver & Stockdale, 2020; Serin et al., 2013; Ward, 2016). As Monahan and Skeem (2016) previously noted, "risk factors known to be causal are in short supply" (p. 506). Although some studies have found that change in certain dynamic risk factors and scales are associated with corresponding change in the perpetration of violent or general recidivism (e.g., Blanchard, 2013; Coupland & Olver, 2018; Penney et al., 2016, Wilson et al., 2013), these findings were often dependent on a number of methodological features, including the dynamic scale of interest, the analysis of focus, the operationalization of the outcome, and other factors included in the analytical models (e.g., de Vries Robbé et al., 2015; Hogan & Olver, 2016, 2019; Michel et al., 2013). As such, there is some empirical evidence supporting the external responsiveness of these instruments, but at the same time other studies have found relatively trivial or no support (e.g., Hogan & Olver, 2019; Mastromanno et al., 2018; Neves et al., 2010). Moreover, only a very limited number of the relevant studies have actually investigated the responsiveness of dynamic risk factors and global ratings of risk state, as opposed to scales comprised of dynamic factors.

Accordingly, some of the purportedly dynamic risk factors included on these tools may tentatively meet the criteria for being labelled as a dynamic or causal risk factor

according to the typologies described above (i.e., Kraemer et al., 1997; Douglas & Skeem, 2005). On the other hand, many of these factors may more appropriately be labelled as variable risk factors or variable risk markers, as there is some evidence supporting their ability to capture intraindividual change over time but no evidence linking this change to change in violence. Some of these factors should simply be labelled by the generic term risk factor, as there is insufficient evidence demonstrating their ability to change over time. However, the labeling of a characteristic with the generic term risk factor is an incomplete and inadequate conclusion, and every conceivable effort should be made to delineate its specific role as a fixed, variable, or causal risk factor (Kraemer et al., 1997). Essentially, "identifying causal risk factors for recidivism is a work in progress that, as yet, cannot support definitive conclusions" (Monahan & Skeem, 2014, p. 161). However, it is possible that through future research using appropriate rigour "today's variable risk factors may become tomorrow's causal risk factors" (Monahan & Skeem, 2014, p. 161).

Based on the available research, it is clear that more work is needed in the identification of true dynamic risk factors (Klepfisz et al., 2016; Serin et al., 2013), that is, risk factors that evidence both internal and external responsiveness. It is a current priority in the field to identify reliable and valid methods of assessing risk that directly and explicitly guide risk management (Ward, 2016). The research to date suggests that purportedly dynamic risk factors are useful for violence prediction and general case formulation, but of lesser or unknown usefulness for monitoring intraindividual change (Serin et al., 2016). Serin and colleagues (2013) noted that "little is confidently known about which factors are the most productive targets for change, whether dynamic risk factors truly behave in a dynamic way to affect future recidivism, and if change predicts recidivism beyond knowledge of static risk" (p. 50). In order to evaluate and improve risk management and rehabilitation efforts, it is necessary to identify the intraindividual factors that are the driving factors behind change in violence and offending. This is a crucial issue to resolve, as the management of violence "hinges on the assumption that positive changes in these risk factors will result in reduced rates of recidivism" (Greiner et al., 2015, p. 459).

Furthermore, almost nothing is known about potential interactions amongst these risk factors over time (Douglas & Skeem, 2005). The explanatory power of dynamic risk factors has been largely unsuccessful (van den Berg et al., 2020). It has been argued

that the limited utility offered by available dynamic risk factors in terms of explaining the causes of violence is partly based on the limited knowledge regarding the relationships amongst these factors (Heffernan & Ward, 2017; Ward & Fortune, 2016).

1.6.1. Methodological and Other Limitations of the Literature

It is not only the case that there is a relative dearth of empirical examinations of change on purportedly dynamic risk factors, but high-quality and methodologically rigorous studies are even rarer (Douglas & Skeem, 2005; Serin et al., 2013). Although several studies have included sound research designs, most empirical studies have suffered from theoretical, methodological, or statistical inadequacies (Brown et al., 2009; Douglas & Skeem, 2005). As well, the available research is still lacking in the number of studies that have employed multiple assessment periods (i.e., more than two) to investigate various types of change patterns and cycles. Much of the research in this area has been subject to methodological limitations in the form of small sample sizes, retrospective and pseudo-prospective research designs, file-based assessments, inconsistent outcome measurement, and too few measurement points to distinguish the nature of change (Douglas & Skeem, 2005). For instance, although Serin and colleagues' (2013) systematic review included 53 studies examining the relationship between change and recidivism, only three of these studies examined violent recidivism and notable methodological limitations were evident in many of the included studies (Klepfisz et al., 2016).

Another limitation of much of the research in this area is the reliance on aggregate data (Klepfisz et al., 2016). The available research has been far too focused on examinations of putatively dynamic risk scales on contemporary risk assessment instruments. Aggregating a number of different dynamic items into a composite may mask significant change occurring on individual risk factors. In addition, risk assessment tools developed under the structured professional judgement model, such as the HCR-20 and START, are not meant to be summed in such a manner. In practice, the communication of risk scale scores (i.e., numerical scores) from these tools is generally avoided, nor is it recommended by their developers. That is, when using a risk assessment tool of this type, the weighting, amalgamation, and prioritization of risk factors is not completed in an actuarial or algorithmic manner but instead determined by the assessment professional. According to current testing standards, it is crucial for

studies to investigate assessment tools as they are intended to be used in practice (AERA, APA, NCME, 2014).

Risk assessment tools developed under the SPJ approach direct evaluators to arrive at a final risk rating based on the totality of the assessment. Accordingly, studies focused on the summation of numerical scale scores may not accurately present the responsiveness of these measures as they were designed and intended to be used (Mills, 2017). As a result, research needs to focus on the individual dynamic risk factors to determine which are capable of change and important characteristics of this change. Research must also focus on whether global ratings of an individual's risk state are capable of change. Research has largely failed to examine changeability in global assessment of risk state made using contemporary risk assessment tools (i.e., final risk judgments made using risk assessment tools developed under the structured professional judgment model). Global ratings of risk state made on these instruments are a defining feature of these tools and require thorough empirical scrutiny.

Another problematic factor is the lack of consensus concerning the best approach for statistically analyzing change data (Brown et al., 2009). A portion of these empirical studies utilized statistical techniques that were based on aggregate data in the form of means or mean ranks (e.g., Belfrage & Douglas, 2002; de Vries Robbé et al., 2015; Douglas et al., 2011; Hogan & Olver, 2016, 2019; Mastromanno et al., 2018; Neves et al., 2010; Olsson et al., 2013). Mean-based tests may conceal intraindividual change. Examining change only at the group level, as opposed to the individual level, has limited applicability to the core feature of whether dynamic risk factors and dynamic scales are sensitive enough to capture intraindividual change over time. For instance, Viljoen and colleagues (2012) reported no mean change in START Strength scale scores across assessments based on a paired-samples t-test, yet according to reliable change indices 11% of Strength scale scores changed over time. Comparatively, a paired sample t-test identified mean change on the Vulnerability scale, yet only 6% of Vulnerability scale scores changed according to the reliable change index analyses. Researchers must select appropriate analysis to address their questions of interest. In this context, it is necessary to analyze change at the idiographic level (Heffernan & Ward, 2017; Kroner & Yessine, 2013), as change may differ from person to person in terms of the speed, frequency, direction, and magnitude.

1.6.2. The Need for Additional Research

There is an ongoing need, in research and practice, to take a dynamic approach to violence risk assessment (Olver & Stockdale, 2020). As Skeem and Monahan called for a decade ago, future research should continue shifting attention from "predicting violence to understanding its causes and preventing its (re)occurrence" (2011, p. 41). From the existing research, it is evident that more research is needed concerning the internal, external, and comparative responsiveness of structured risk assessment instruments (Beech et al., 2016; Brown et al., 2009; Cording et al., 2016; Mann et al., 2010). That is, research should continue to explore the extent to which dynamic risk factors on structured risk assessment tools in fact fluctuate over time and the extent to which this change is associated with meaningful change in violence. Due to the nature of responsiveness, it is also necessary to place more emphasis on validating and replicating prior research findings to identify "true phenomena rather than idiosyncrasies of particular datasets" (Cording et al., 2016, p. 91). It is hoped that further research will provide insights into how these factors independently and collectively generate and maintain violence.

Assessment and management of violence risk should concentrate on empirically based causal risk factors (Mann et al., 2010). Further research is required to empirically validate dynamic (causal) risk factors and the extent to which change on these factors leads to reduction in violent recidivism (Mann et al., 2010). In order to advance the effectiveness of prevention and intervention efforts, the field needs a better understanding of the specific role of purported dynamic risk factors in leading to violence (van den Berg et al., 2020). Specifically, research is needed regarding the nature and strength of the direct, indirect, and cumulative associations between dynamic risk factors and violence, as well as the interrelationships amongst these various risk factors (Klepfisz et al., 2016; van den Berg et al., 2020). Future research should also examine the independent and collective extent to which change on dynamic risk factors and dynamic scales impacts global assessments of risk (Klepfisz et al., 2016).

Forthcoming research should ideally employ a multiple time-point, longitudinal, prospective research design (with a minimum of three waves) and feature a breath of potentially dynamic risk factors and frequent measurements of violence (Brown et al., 2009; Douglas & Skeem, 2005; Jones et al., 2010). These features will provide enough

measurements to discern different trajectories or patterns of change, based on the frequency and speed of change in the risk factors, as well as the association between this change and violence. As well, research should focus on commonly used structured risk assessment tools that are comprised of a set of hypothetically dynamic risk factors (Klepfisz et al., 2016). These recommendations should be coupled with statistical techniques that can address intraindividual change over time (Brown et al., 2009). Douglas and Skeem (2005) also called for research focusing on the interactive, moderating, and mediating effects of one variable on another, and the incremental predictive power of dynamic risk factors and change on these factors over and above static risk factors.

1.7. Purpose of the Current Research

Based on this review of the relevant theoretical and empirical literature, the current research was undertaken to build on the extant empirical literature. Using a longitudinal, multi-wave research design, the present study aimed to investigate the internal, external, and comparative responsiveness of putatively dynamic risk factors and comprehensive ratings of risk state using the HCR-20 and START. This study was undertaken to add knowledge to the field concerning the extent to which these risk assessment tools are sensitive to change over time and the relationship between this change and the propensity for violence.

Overall, the study focused on two main branches of investigation. First, research questions were formulated in an attempt to empirically identify dynamic risk factors. Generally, these research questions focused on identifying whether putatively dynamic risk factors are capable of change across assessments and whether this change is associated with change in the likelihood of violence. Second, research questions addressed the ability of these contemporary risk assessment tools to assess an individual's risk state. In general, these research questions focused on identifying whether summary risk judgments are capable of change across assessments and whether this change is associated with change in the likelihood of violence.

Specifically, the current study aimed to answer the following research questions regarding the responsiveness of dynamic risk ratings on risk assessment protocols (i.e., ratings of putatively dynamic risk factors, dynamic scales, and overall risk judgments):

1. Do dynamic risk ratings change over multiple assessments? This question concerned the changeability of dynamic risk ratings. As a defining feature, dynamic ratings must be capable of change over time. This question ultimately attempted to address the internal responsiveness of these ratings.

    A. Are there identifiable patterns or associations in change? If change is seen across assessments, this inquiry attempted to identify the interrelationships between different dynamic ratings. It may be the case that change on one dynamic risk factor influences change on others (e.g., change in Active Symptoms of Major Mental Illness impacting change on Lack of Insight and/or Noncompliance with Treatments). This question examined an important feature of internal responsiveness.

2. Does change on dynamic risk ratings relate to change in violence? This research question addressed a primary requirement of dynamic risk ratings; change must relate to change on the outcome. This question addressed the external responsiveness of these dynamic risk ratings.

3. How do the various dynamic ratings compare within and across tools with respect to their responsiveness? This question focused on differences in the internal responsiveness within and across the two risk assessment tools, and thus addressed comparative responsiveness.

Chapter 2. Method

To assure the transparent reporting of important methodological and statistical details and study results, this dissertation adheres to the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement (Singh et al., 2015), a 50-item reporting checklist.[2]

2.1. Study Participants

The current project combined data from a sample of correctional offenders (*n* = 86) and a sample of civil psychiatric inpatients (*n* = 149) for a total sample size of 235 at the baseline phase.[3] Of this total sample, 216 participants were assessed on the relevant risk assessment instruments. The sample size for each of the five follow-up phases is presented in the Sample Size and Attrition section below (see Table 3-1). The characteristics of each of the subsamples at the baseline phase are included in turn.

2.1.1. Civil psychiatric sample

A total of 149 civil psychiatric inpatients were recruited from the acute stay ward of a large hospital in the greater Vancouver metropolitan area in British Columbia (i.e., Royal Columbian Hospital). Both voluntarily and involuntarily admitted patients were included in the study. To be referred for inclusion in the study, participants needed to be (a) between the ages of 19 and 50,[4] (b) able to speak in English, (c) planning to reside in the metropolitan area, and (d) not diagnosed with any severe intellectual disability. Staff psychiatrists on the inpatient ward referred eligible patients to the research team. Subsequently, a member of the research team spoke with the patient to describe the study, answer any questions, and guide them through the consent procedures. All

[2] Disclosure: The senior supervisor (K.S.D) and one committee member (S.D.H.) are authors of the HCR-20.

[3] Disclosure: The data for this study was collected as part of two larger longitudinal studies of which the senior supervisor was principal investigator; this data has been used in several previous publications, dissertations, theses, and presentations, including Blanchard (2013).

[4] This restriction on age was relaxed during data collection. A total of 5 participants over the age of 50 were included in this subsample, resulting in an age range from 19 to 61 years old.

patients were assessed for their competence to participate in the study using a multiple-choice test included in the consent procedures.

The civil psychiatric sample was roughly evenly split across gender (51.7% men). The majority of the sample was White (79.2%) with the next largest group identifying as being of Asian descent (10.7%). As well, a small proportion of the sample identified as Indigenous (2.7%), Black (0.7%) or some other ethnicity (6.0%). The average age across participants was 33.80 years (*SD* = 10.18). Participants in the civil psychiatric sample had an average of 12.44 years of education (*SD* = 2.14), with 28.1% of the sample never completing high school and 40.0% having completed at least some college level courses.

Most participants had never been married (62.4%), with smaller proportions being currently married (18.1%) or separated or divorced (18.1%). Relatedly, most of the sample was not currently involved in a long-term romantic relationship (75.7%), nor did they have any children (67.6%). Prior to being admitted to the hospital most participants were either living with family (36.2%) or living alone (30.9%), with smaller proportions living in subsidized housing units (14.8%), having no fixed address (4.7%), or some other living situation (9.4%). A notable minority of the participants (32.1%) had experienced a time in which they were homeless (i.e., no fixed address) prior to the baseline phase.

Most patients were admitted to the psychiatric ward involuntarily (77.2%). Over the course of their lifetime prior to the baseline assessment phase, the most common mental health diagnosis was a psychotic disorder (44.9%), followed by bipolar disorder (39.5%), any type of depressive disorder (38.1%), any type of substance disorder (33.3%), or any type of anxiety disorder (21.1%). This was not the first psychiatric hospitalizations for most patients; 68.7% of the sample had at least one prior hospitalization with an average of 2.65 (*SD* = 3.04) prior hospitalizations. A small but noteworthy proportion of the sample had previously been incarcerated (17.8%) with an even larger proportion coming into formal contact with the criminal justice system (i.e., arrest, charge, or conviction) for a non-violent offence (36.2%) and a smaller proportion coming into formal contact with the justice system for a violent offence (15.9%).

2.1.2. Correctional sample

A total of 86 correctional offenders were recruited from four correctional institutions (i.e., Fraser Regional Correctional Centre, North Fraser Pretrial Centre, Surrey Pretrial Services Centre, and Allouette Correction Centre for Women) and four probation offices (i.e., Burnaby Probation & Family Court, Abbotsford Community Corrections, Tri-Cities Community Corrections, and Vancouver Intensive Supervision Unit) in the greater Vancouver metropolitan area in British Columbia. Offenders must have been sentenced to the custody of a provincial corrections institution or provincial probation office in the catchment area to be eligible for inclusion in the study. To be under the care of provincial corrections, offenders must have been sentenced to a maximum custody term of two years minus a day. Both mentally disordered and non-mentally disordered offenders were included, with mentally disordered offenders defined as those with a lifetime history of any major mood or psychotic disorder.

Incarcerated offenders must have been serving a minimum of one month in custody prior to being referred and they must have been within one month of release into the community. Offenders serving community sentences (i.e., probation and conditional sentences) must have been under the care of provincial corrections for at least one month prior to be referred and they must have been under supervision for at least six additional months. In addition to these requirements, the same eligibility criteria were used as the civil psychiatric sample (i.e., age,[5] location, and language requirements, and lack of severe intellectual disability). Recruitment procedures were similar to the civil psychiatric sample. Correctional staff from the institutions referred eligible offenders to the research team. Some correctional sites also advertised the project directly to offenders who then self-selected for a referral to the study. A member of the research team then spoke with the offender to describe the study, answer any questions, and guide them through the consent procedures. All offenders were assessed for their competence to participate in the study using a multiple-choice test included in the consent procedures.

[5] The inclusion criteria requiring offenders to be between the ages of 19 and 50 was similarly relaxed during data collection. A total of 2 participants over the age of 50 were included in this subsample, resulting in an age range from 19 to 61 years old.

The correctional sample was also roughly evenly split across gender (51.2% men). The majority of the sample was White (73.8%) with the next largest group identifying as Indigenous (15.5%). As well, a small proportion of the sample identified as Asian (1.2%) or some other ethnicity (6.0%). The average age across participants was 34.65 years (*SD* = 8.41). Participants in the correctional sample had an average of 10.94 years of education (*SD* = 1.70), with 46.7% of the sample never completing high school and 8.3% having completed at least some college level courses.

Most participants had never been married (64.7%) with smaller proportions being currently married (12.9%) or separated or divorced (20.0%). Relatedly, most of the sample was not currently involved in a long-term romantic relationship (62.7%), nor did they have any children (54.7%). Prior to their sentence, most participants were living on the street (25.6%), living alone (23.2%), or living in subsidized housing units (17.1%), with smaller proportions living with family (12.2%) or with friends (12.2%). Most participants (75.9%) had experienced a time in which they were homeless prior to the baseline phase.

The sample included both mentally disordered offenders (60.5%) and non-mentally disordered offenders (39.5%). Over the course of their lifetime prior to the baseline assessment phase, the most common mental health diagnosis was a substance use related disorder (59.4%), followed by any type of depressive disorder (34.4%), an anxiety disorder (31.3%), bipolar disorder (25.0%), or a psychotic disorder (21.9%). A notable portion of the sample (44.0%) had been hospitalized for a mental health reason prior to the baseline phase, with an average of 1.49 (*SD* = 2.86) psychiatric hospitalizations.

The correctional sample was roughly equally split between offenders serving custodial sentences (54.5%) and offenders serving community sentences (45.5%). The most common sentence included a term of incarceration followed by a term of community probation (49.4%); other offenders were sentenced to only community probation (30.1%), or only a term of incarceration (18.1%). Most of the participants had a previous provincial incarceration sentence (67.5%), and a small proportion had a previous federal incarceration sentence (14.6%). The most common index offences were crimes against property (57.0%), specifically theft (32.6%). Crimes against persons were

also common (40.7%), including various forms of assault (19.8%). Drug offences were less common (11.6%), including trafficking offences (7.0%).

2.1.3. Sample Comparisons

The two subsamples were compared across basic demographic variables. No differences were observed in age, gender, marital status, or number of children. Some differences were observed in the proportions of each ethnicity represented in the two samples, χ^2 (5, N = 233) = 22.16, p = <.001, with a greater number of Indigenous participants in the correctional sample and a greater number of Asian participants in the psychiatric sample. As well, the civil psychiatric sample had more years of education on average (M = 12.79, SD = 2.30) than the correctional offenders (M = 11.22, SD = 2.06), t(219) = 5.11, p < .001, d = 0.71, 95% CI = [0.43, 0.99]. There were no differences observed in the history of various mental health disorders, except for a history of a psychotic disorder diagnosis, χ^2 (1, N = 179) = 5.77, p = .016, with more participants in the psychiatric sample with a history of psychotic diagnosis (44.9%) compared to the correctional sample (21.9%). Prior psychiatric hospitalizations were more common in the psychiatric sample (68.7%) than the correctional sample (44.1%), χ^2 (1, N = 231) = 13.51, p <.001, and the psychiatric sample had a greater number of prior psychiatric hospitalizations (Mdn = 2.00 to 0.00), U = 3956, Z = -4.03, p < .001. In addition, the correctional offenders (M = 12.73, SD = 4.39) had a higher average psychopathy scores than the psychiatric patients (M = 6.66, SD = 4.88), t(212) = -9.08, p < .001, d = -1.29, 95% CI = [-1.59, -1.03].

2.2. Measures

2.2.1. Historical-Clinical-Risk Management-20 (HCR-20).

The HCR-20 (Webster et al., 1997) is one of the original violence risk assessment instruments developed under the Structured Professional Judgment (SPJ) model of risk assessment. The current study used version two of the HCR-20, as version three (Douglas, Hart, et al., 2013) was still under development and not implemented at the time that data collection began. The HCR-20 has been identified as the most used violence risk assessment instrument among forensic mental health professionals across the globe (Arai et al., 2017; Singh et al., 2014; Ramesh et al., 2018).

It is designed to aid professionals in conducting a violence risk assessment on an adult from a variety of settings and populations (Douglas, Hart, et al., 2013). It was created to assist in violence risk determinations for civil psychiatric, forensic psychiatric, and criminal offenders (mentally disordered or not). It is frequently used in a multitude of decision contexts, including bail and sentencing decisions, admission and release planning, and monitoring treatment and other progress (Douglas & Shaffer, 2020). It is recommended that information from five general sources be used to complete the HCR-20: file or official record information, an interview with the individual undergoing the assessment, psychological or other test data, direct observation, and interviews with relevant collateral sources (Douglas, Hart, et al., 2013; Douglas & Shaffer, 2020).

The HCR-20 is comprised of 20 risk factors organized into three domains or scales: the Historical scale, the Clinical scale, and the Risk Management scale. The measure contains both static items on a single scale (i.e., Historical) and purportedly dynamic items on two scales (i.e., Clinical and Risk Management). Specifically, it includes 10 Historical items related to the individual's past functioning, five Clinical items related to the individual's recent functioning, and five Risk Management items related to the individual's future functioning (see Table A1). The Risk Management items reflect the professional's opinions about the individual's ability to adjust to either an institution (i.e., "in" ratings) or community (i.e., "out" ratings) setting. For the current study, the "out" ratings were used as the project focused on risk in the community.

Each risk factor is rated on a three-point system, a coding practice common amongst risk assessment tools developed under the SPJ model. Risk factors coded as Absent (No or 0) indicate that the specific feature is absent or does not apply to this individual. Risk factors coded as Possibly Present (Possible or 1) indicate that the feature is present only to a limited degree or the individual's manifestation of this feature only partially matches the item definition. Risk factors coded as Present (Yes or 2) indicate that the feature is or was present during the relevant timeframe. As a SPJ tool, the individual items are not meant to be scored numerically, nor are they intended to be summed or aggregated into an overall numerical score in clinical practice. However, numeric scale and total scores are typically used in research contexts to evaluate the tools psychometric properties (Douglas, Hart, et al., 2013; Douglas et al., 2014). Notably, one of the items on the Historical scale is based on the level of psychopathic traits and requires the completion of either the *Psychopathy Checklist – Revised* (PCL-R; Hare,

1991; which was used with the correctional sample) or the *Psychopathy Checklist – Screening Version* (PCL-SV; Hart et al., 1995; which was used with the civil psychiatric sample).

After coding each of the individual risk factors, a professional using this tool is required to summarize their overall opinions on another three-point system: Low, Moderate, or High risk. To arrive at these final opinions, professionals should incorporate all information about the case, including the number of risk factors that are present, the idiographic relevance of these risk factors to the given case, and the types and intensity of risk management services that are required to mitigate this individual's risk. These global ratings are meant to indicate the individual current risk state (i.e., a potentially dynamic rating of the individual's current risk).

The HCR-20 has been the subject of extensive empirical investigations and is possibly the most well researched violence risk assessment protocol in existence. Hundreds of examinations of this tool's psychometric properties have been undertaken (e.g., Douglas & Reeves, 2010; Douglas & Shaffer, 2020; Douglas et al., 2014; Guy, 2008). The existing research has largely concluded that the HCR-20 has good to excellent predictive validity and interrater reliability. For instance, studies have shown the HCR-20 to be predictive of violence in community (e.g., de Vogel et al., 2004; Gray et al., 2003; Pedersen et al., 2010; Penney et al., 2016) and institutional settings (e.g., Arai et al., 2017; Belfrage et al., 2000; Hogan & Olver, 2016; Ramesh et al., 2018).

Research on the HCR-20 has been subjected to both narrative (e.g., Douglas et al., 2014; Douglas & Shaffer, 2020) and meta-analytical (e.g., Campbell et al., 2009; Guy, 2008; Ramesh et al., 2018; Yang et al., 2010) reviews confirming the results of the large body of independent studies. With respect to interrater reliability, narrative reviews have reported reliability coefficients in the good to excellent range (i.e., ICC = 0.67 to 0.95), with most studies finding reliability coefficients greater than 0.80 (Douglas & Reeves, 2010). With respect to predictive validity, a meta-analysis of violence risk assessment studies determined that the HCR-20 numerical scores (mean AUC = 0.73) and summary risk judgments (mean AUC = 0.76) were predictive of violence (Guy, 2008). O'Shea and colleagues (2013) also reported the largest predictive effects for the summary risk judgments compared to numerical scores; moreover, the dynamic scales consistently outperformed the static scale in predicting inpatient violence.

Other meta-analyses have reported that on average the HCR-20 yielded effect sizes that were at least as large, if not larger, as those yielded by other contemporary risk assessment measures (e.g., Campbell et al., 2009; Yang et al., 2010). Based on a meta-analysis of 78 unique samples including 6,840 psychiatric patients, Ramesh and colleagues (2018) found the HCR-20 was moderately predictive of inpatient violence (median AUC = 0.70). Similarly, Hogan and Ennis (2010) found support for the predictive validity of inpatient violence in their meta-analysis (mean weighted r = 0.33). Another meta-analysis based on 45 effect sizes found support for the prediction of violence using HCR-20 in samples of women (Rossdale et al., 2020).

2.2.2. Short-Term Assessment of Risk and Treatability (START).

The START (Webster et al., 2009) is a newer risk assessment instrument developed under the SPJ model that is designed to assess risk across seven domains: violence to others, suicide, self-harm, victimization, substance use, unauthorized absence, and self-neglect. Although the START is intended to facilitate the assessment of all seven outcomes, the current study only focused on the risk of violence to others. The START is also one of the most commonly used structured risk assessment instruments across the globe (Singh et al., 2014).

This tool is designed to assist with the assessment of risk on an adult from a variety of settings and populations (Desmarais et al., 2012). Specifically, it is intended for use with civil psychiatric, forensic psychiatric, and criminal offenders in both inpatient and outpatient settings. It is intended to be used by interdisciplinary professionals to inform judgments about meaningful risks and strengths, as well as guide treatment decisions (Webster et al., 2006; Nicholls et al., 2006). The START is meant to facilitate the communication, planning, and management of the above-mentioned risks, particularly in the context of working with interdisciplinary treatment teams. Overall, this specialized instrument focuses on dynamic risk and multiple risk domains that are relevant to daily psychiatric practice. It encourages professionals to concentrate on the individual's current and changing circumstances to guide intervention plans to minimize any risks and improve strengths.

The START is comprised of 20 dynamic factors that include both positive (i.e., Strengths) and negative (i.e., Vulnerabilities) descriptions. That is, this tool is comprised

of two domains or scales, and each of the 20 dynamic items receives a rating as both a Strength and a Vulnerability (see Table A2). Of note, the rating of Strengths and Vulnerabilities are considered independent; a person may have any combination of Strength and Vulnerability ratings on single item.

As an SPJ instrument like the HCR-20, each item is rated on a three-point system with the same general meaning associated with each rating. That is, each Strength and Vulnerability item is coded independently as Absent (No or 0), Possibly Present (Possible or 1), or Present (Yes or 2). Again, like the HCR-20, the individual items are not meant to be scored numerically, nor are they intended to be summed or aggregated into an overall numerical score in clinical practice (Webster et al., 2009). However, numeric scale scores are typically used in research contexts to evaluate the tools psychometric properties (Hogan & Olver, 2018; Wilson et al., 2013). With respect to interpreting scale scores, higher Vulnerability scale scores indicate greater risk, whereas higher Strength scale scores indicate lower risk.

After coding each of the individual Strength and Vulnerability factors, professionals are required to summarize their overall opinions concerning each of the seven domains on a three-point system: Low, Moderate, or High risk. Again, these global risk ratings are not made based on normative data, an algorithm, or other mechanical means. In order to arrive at these final opinions, professionals should incorporate all information about the case, including the number and combination of Strength and Vulnerability factors that are present, the idiographic relevance of these factors to the given case, the identification of key and critical items (i.e., items of particular significance to this individual's risk), the number and relevance of historical factors that are present (possibly assessed using the HCR-20 Historical scale), and the types and intensity of risk management services that are required to mitigate this individual's risk profile. These global ratings indicate the individual's current risk state.

Although the START is a comparatively new instrument and has not been subject to the same extent of empirical investigation as the HCR-20, there is a growing body of literature on the properties of the START (O'Shea & Dickens, 2014; Whittington et al., 2014). Dozens of studies have examined this tool's psychometric properties (e.g., Braithwaite et al., 2010; Chu et al., 2011; Desmarais et al., 2012; Doyle et al., 2008; Hogan & Olver, 2016, 2019; Gray et al., 2011; Nicholls et al., 2006; Nonstad et al., 2010;

Wilson et al., 2013). Overall, it may be largely concluded from these studies that the START has good to excellent predictive validity and interrater reliability with respect to predictions of violence. For instance, Nicholls and colleagues (2006) examined the performance of the START in a sample of 137 forensic psychiatric inpatients. Looking at the total numerical scores, the authors reported excellent interrater reliability (ICC = 0.87) and good predictive validity with regards to violence to others (AUC = 0.70). Similarly, Wilson and colleagues (2013) reported excellent interrater reliability for the Strength (ICC = 0.85) and Vulnerability (ICC = 0.90) scales, as well as the summary risk judgments (ICC = 0.81). The START scales and summary risk judgments were also found to be predictive of violence over a 12-month follow-up (AUCs = 0.82 to 0.89). As well, Desmarais et al. (2012) reported excellent interrater reliability, as well as both predictive and incremental validity over the HCR-20 Historical scale scores.

Research on the START has also been subjected to both narrative (e.g., Chu et al., 2011; Nicholls et al., 2021) and meta-analytic (e.g., O'Shea & Dickens, 2014) reviews that have corroborated the results of the independent studies. For instance, the results of seven studies that have examined the predictive validity of the START for inpatient violence have been summarized by Chu et al. (2011). Overall, evidence was found for the predictive validity of the Vulnerability scale (AUCs = 0.63 to 0.83), Strength scale (AUCs = 0.73 to 0.77), and summary risk judgments (AUCs = 0.52 to 0.82) for the prediction of any violence. In addition, O'Shea and Dickens (2014) integrated the results of 23 studies that examined the properties of the START in a narrative review and further analyzed the results of nine studies based on 543 participants in a meta-analysis. In general, START scores demonstrated high interrater reliability and convergent validity with other risk assessment protocols. The START scales and summary risk judgments were also moderately predictive of various indices of violence (AUCs = 0.71 to 0.74 for any aggression). One the other hand, there is a relative lack of examinations into the other outcome domains.

2.2.3. Definition and Measurement of Violence.

The current study used the definition of violence originally provided by the HCR-20 (Webster et al., 1997). That is, violence is any "actual, attempted, or threatened harm to a person or persons" (p. 24). With respect to the measurement of violence, two distinct methods or sources were used: a semi-structured interview with participants (i.e.,

self-reported violence) and file information (i.e., correctional records were available for the correctional sample, and community mental health records were available for the civil psychiatric sample).

In the interview protocols, violence was documented using the MacArthur Community Violence Interview (Monahan et al., 2001). This interview system queries regarding nine diverse categories or types of violent behaviour ranging from minor violent acts (i.e., throwing something at another person) to potentially more harmful acts (i.e., pushing, hitting, choking) all the way to life-threatening acts (i.e., using a gun or other weapon on another person). This interview protocol also includes an "other" category to capture additional violent behaviours that may not be appropriately captured by the prior eight types of violence. In addition to the MacArthur Community Violence Interview, some supplementary types of violence were also measured in the larger interview to align with the definition of violence from the HCR-20 more accurately. The supplementary types of violence included acts such as threats of violence without a weapon and intentionally causing meaningful fear in another person.

Violence was collapsed in the present study into two dichotomous outcomes of interest. The broad or overall operationalization of violence included all the aforementioned types or categories of violence from both sources of information. In contrast, the narrow or serious operationalization of violence included only the more serious or potentially harmful forms of violence (i.e., only those categories included on the MacArthur Community Violence Interview).

2.3. Procedure

2.3.1. Overall study design and assessment schedule.

This study employed a prospective, longitudinal, repeated measures research design. Participants were assessed at baseline and then approximately once a month for up to five follow-up assessments. Thus, each participant was subject to up to six assessments. All the assessments included a semi-structured interview, a review of relevant collateral information (e.g., correctional or mental health files), and completion of a self-report test battery. Some assessments also included interviews with close collateral contacts (e.g., spouse, parent, sibling, friend, etc.). The dynamic items on the

HCR-20 and the START protocol were completed at each of the assessments. As well, violence was measured at each of the assessments. For the civil psychiatric sample, the baseline assessments took place during the month prior to being discharged or shortly after being discharged from the psychiatric ward. For the correctional sample, the baseline assessment took place during the month prior to being released for those offenders serving provincial prison sentences and anytime during their probation term for those offenders serving community probation sentences.

The participant interviews for both samples queried about a broad array of information relevant to rating the HCR-20 and START protocols, including questions related to violence perpetration, mental health problems, substance and alcohol use, social support and relationships, treatments and other professional services, a variety of cognitions and attitudes, and plans for the future. The collateral interviews similarly inquired about a host of attitudes, cognitions, and behaviours relevant to detecting violence and rating the risk assessment instruments. For the civil psychiatric sample only, the baseline assessments included a review of their hospital records up to and including the index hospitalization. During the follow-up assessments for those participants attending a select number of outpatient community mental health clinics, the assessments included a review of their mental health outpatient files. For the correctional sample only, the baseline and follow-up assessments included a review of their correctional files (if the participant was still under the supervision).

2.3.2. Risk assessment professionals.

The interviews and risk assessment protocols were completed by trained research personnel. The researchers were comprised of students ranging in experience from senior undergraduate students to senior doctoral students. Most researchers were graduate students enrolled in a clinical or experimental forensic psychology program. All the risk assessment personnel underwent specialized training for the current project in interviewing procedures and skills, general assessment procedures, and completion of the two risk assessment instruments.

With respect to training in information gathering practices, the researchers were trained to conduct the interviews in a professional and standardized manner; this included training on the specific semi-structured interview used in the current study,

observing an interview conducted by a trained professional, and being observed during their first interview. Periodic refresher training was also provided throughout the project.

With respect to training on the two risk assessment protocols, the researchers received specialized training in completing the respective protocol from an author of each of the risk assessment measures. The risk assessment training included the completion of assigned readings, a thorough review of the instrument's manual, the completion of several standard practice cases based on video recorded interviews and case vignettes, and feedback from the trainer concerning the practice cases. As well, multiple researchers completed the risk assessment protocols on preliminary cases until the researchers reached acceptable levels of interrater reliability.

2.4. Statistical Analyses

Descriptive statistics for all variables were computed and compared between the civil psychiatric and correctional samples, using chi-squared tests, t-tests, and Mann-Whitney tests as appropriate.[6] Subsequently, as this study was one of the first to examine intraindividual change in dynamic risk factors, the two subsamples were combined. The samples were combined to increase the number of observations included in each analysis and consequently improve the statistical power of these tests.

Based on the research questions of interest, the main inquiries centre on the capacity of these instruments to detect change overtime and not on the performance of the instruments within each sample. Moreover, responsiveness is not considered a stable property of an instrument; instead, it is dependent upon several methodological and statistical characteristics of a given study (Beaton et al., 1997, 2001; Freeman et al., 2013; Husted et al., 2000).

2.4.1. Missing data.

Missing data was addressed in several ways. In the current project, missing data can be categorized based on the amount of and reasons for the missing data. Most of the missing data was completely missing for an entire assessment. That is, completely

[6] Note, the significance criterion for these and all other hypothesis tests was set at 0.05. See the Limitations and Qualifications section for a discussion of familywise error.

missing data resulted from participants leaving the study prior to completion (i.e., attrition), in which the participant failed to complete a follow-up assessment and all subsequent follow-up assessments.[7] A smaller proportion of data was partially missing within each assessment. Partially missing data arose when participants completed the assessment, but certain observations or variables were missing (e.g., omitted items, lack of ratings, or other unknown errors).

Participants with completely missing data and partially missing data on individual ratings included on the violence risk assessment protocols (i.e., the HCR-20 and START) were excluded from the relevant item level analyses (i.e., pairwise deletion) or a procedure was used that can accommodate different numbers of observations per participant. That is, in circumstances when participants were missing data for an entire assessment period or missing item level ratings, the missing data was excluded from the relevant comparisons (i.e., RCI, Wilcoxon, and Spearman's) or the procedure could handle varying numbers of measurements per participant (i.e., repeated measures correlations and generalized estimating equations).

The current study did not select to employ multiple imputation due to the nature of the study and the requirements of the statistical techniques. As responsiveness is dependent upon several methodological factors (e.g., Freeman et al., 2013; Husted et al., 2000), it is not the intention of the current study to generalize the specific findings or characteristics of responsiveness to other contexts or populations. Instead, this study was undertaken to determine the extent to which these tools may be capable of capturing change and describe some of the various properties of change that may be important for tools of this manner. Moreover, multiple imputation was considered unsuitable in the current context, as it may result in increased error terms and alter statistical coefficients (e.g., Enders, 2011; Garson, 2015; Grittner et al., 2011; Newman, 2003; Rubin, 1987). These effects are potentially magnified when there is a high proportion of missing data and dependent upon the particulars of the statistical analysis. Multiple imputation combined with analytical techniques investigating change over multiple assessments may in fact mask or delude changes in scores.

[7] For details regarding the number of participants assessed in each follow-up period and corresponding missing data, see the Sample Size and Attrition section below.

Participants with partially missing data on the violence risk assessment protocols (i.e., the HCR-20 and START) were addressed by prorating the scale and total scores using instructions provided in the respective manuals. With regards to the HCR-20, scores on the static scale (i.e., Historical) were prorated if no more than two items were missing, whereas scores on the two dynamic scales (i.e., Clinical and Risk Management) were prorated if no more than one item was missing from a given scale. At the baseline assessment, a single participant was missing two Clinical items and all five Risk Management items and was therefore excluded from scale level analyses.[8] Throughout the follow-up phase, only a single assessment from the second follow-up contained an omitted item on the Clinical scale and none contained omitted items on the Risk Management Scale. No data was missing for the HCR-20 summary risk judgments.

With regards to the START, scores on the two scales were prorated if no more than 4 items were missing from a given scale. To prorate these scales, the average item score for the available items on the relevant scale was multiplied by the total number of items on that scale. At the baseline assessment, six participants were missing a single item on the Vulnerability scale, while eight participants were missing a single item and one participant was missing two items on the Strength scale. An additional two participants were missing too many items on each of the two scales to be able to prorate their scale scores. Two participants were missing the summary risk judgments for violence at baseline. Throughout the follow-up phase, eight assessments were missing a single item on the Vulnerability scale, while three assessments were missing two items. Additionally, 11 assessments were missing a single item on the Strength scale, while an additional two assessments were missing two items on this scale. Three assessments were missing the summary risk judgments from the START during the follow-up phase.

2.4.2. Interrater reliability.

Interrater reliability on the violence risk assessment protocols was only examined in the civil psychiatric sample. Specifically, interrater reliability was addressed by means of two independent researchers coding approximately 25% of the baseline assessments

[8] Note, the number of missing ratings described in this paragraph and the subsequent paragraph refer only to partially missing data (i.e., missing data from completed assessments). Completely missing data (i.e., no data collected for one or more entire assessments) is described in the Sample Size and Attrition section (see Table 3-1).

in this sample only (*n* = 31). Both researchers attended a single interview with the participant and conducted separate file reviews to independently complete the risk assessment tools for these cases.

Intraclass correlation coefficients (ICCs) were employed to statistically appraise the level of interrater reliability. In particular, the current study employed a two-way random effects model appraising absolute agreement. ICCs are optimally suited for such tasks, as they are mathematically equivalent to a weighted kappa and provide a chance corrected measure of agreement rather than simply reporting on the strength of the association (such as Pearson's r) between the two ratings (e.g., Bartko & Carpenter, 1976; Cicchetti & Sparrow, 1981; Douglas et al., 2003; Landis & Koch, 1977). The ICCs were interpreted according to the categorical descriptions provided by several prominent sources. That is, in describing kappa and weighted kappa coefficients, Cicchetti and Sparrow (1981) defined reliability indices from .00 to .40 as poor, .40 to .59 as fair, .60 to .74 as good, and above .75 as excellent. On the other hand, Landis and Koch (1977) defined reliability indices from .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.

2.4.3. Reliability of measurement.

Several techniques were employed to examine the reliability of measuring change on the risk assessment protocols. The Standard Error of Measurement (SEM) was calculated based on the reliability and standard deviation of the respective portion of each of the risk tools (Jacobson & Truax, 1991; Stratford et al., 1996). Specifically, the SEM was calculated using the relevant standard deviations from all assessments using the combined samples, as well as the reliability of the tools measured at baseline. Interrater reliability was used in computing the SEM as this form of reliability is considered most important for measures of this sort and this approach is consistent with previous research in the area (e.g., Draycott et al., 2012; Viljoen, Shaffer, et al., 2017). The SEM was then used to estimate the minimally detectable change (MDC), or error threshold, which indicates the smallest true difference (i.e., change) that an instrument is capable of detecting (Beckerman et al., 1996, 2001; Schuck & Zwingmann, 2003; Stratford et al., 1996). In addition, floor and ceiling effects are reported, indicating the number and proportion of participants that were unable to demonstrate a decrease or increase in scores due to already having the minimum or maximum score, respectively.

2.4.4. Internal responsiveness.

Several approaches were undertaken to examine the internal responsiveness (i.e., changeability) of the risk assessment tools. Fundamentally, change scores were calculated for ratings across each assessment pair (i.e., baseline score minus follow-up one score, follow-up one score minus follow-up two score, etc.). As such, for the risk factors and risk scales positive scores indicate improvement (i.e., decrease in the presence of risk factors), while negative scores indicate worsening (i.e., increase in the presence of risk factors). In contrast, for the protective factors and scale (i.e., START Strength scale), positive scores indicate a worsening (i.e., decrease in the presence of protective factors), while negative scores indicate an improvement (i.e., increase in the presence of protective factors).

Two general approaches were used to examine intraindividual change on the risk assessment protocols. First, the proportion of participants who evidenced change on the risk assessment tools was examined for each type of rating on the tools (i.e., items, scales, and summary risk judgements), including the number of participants whose scores changed between any of the six assessments and the number of participants whose scores changed between any two assessment points. Second, a reliable change index (RCI) was calculated for each pair of assessments using the approach described by Jacobson and Truax (1991). Under this approach, the RCI corresponds to an individual's change score divided by the standard error of the difference. The standard error of the difference was computed from the SEM.[9] RCI is an extension of the MDC that assists the interpretation of change for a specific pair of scores (Viljoen, Shaffer, et al., 2017). That is, RCIs facilitate the examination of intraindividual change by allowing for the comparison between a reliable change index and a critical value based on the standard normal distribution (Draycott et al., 2012).

The RCI presents a Z-value that signifies the probability that change is observed due to measurement error or chance alone, as opposed to true change (Evans et al., 1998; Hinton-Bayre, 2010; Riddle & Stratford, 2013). This technique provides a standardized score that can be compared to the critical value of 1.96, which corresponds to an alpha of 0.05 in conventional parametric testing. Therefore, if a given RCI is

[9] The standard error of the difference was computed as $s_{diff} = \sqrt{[2(SEM)^2]}$.

greater than 1.96, it indicates that the probability that this change is due to chance alone is less than 5% (Viljoen, Shaffer, et al., 2017). This study used the approach developed by Jacobson and Truax (1991) based on its computational simplicity and intended application. Other approaches are meant to account for practice effects and regression to the mean, and studies have generally failed to find substantial differences between the various methods available (McGlinchey et al., 2002; McGlinchey & Jacobson, 1999; Speer & Greenbaum, 1995; Wise, 2004).

For some of these analyses, each participant's assessments were collapsed across the entire sample to examine the total number of reassessments. That is, these analyses focused on the total number of assessment pairs across all participants. Thus, each assessment pair (i.e., baseline to follow-up one, follow-up one to follow-up two, etc.) was included as a separate case, allowing for each participant to contribute a minimum of one and a maximum of five observations. For these specific analyses, collapsing across assessments effectively changed the overall methodology from a multiple time-point design to a dual time-point design.

In addition, two general approaches were also used to examine aggregate-level change on the risk assessment protocols. Although the current study is focused on intraindividual change, aggregate-level change analyses have their value. Though the absence of aggregate-level change does not indicate the absence of intraindividual change, the presence of aggregate-level change signifies the presence of some intraindividual change. As well, aggregate-level analyses allow for comparisons across type of analysis (intraindividual versus aggregate) and afford an opportunity to directly compare with prior research using similar techniques. As such, Wilcoxon signed-rank tests and Spearman's rank-order stability coefficients were used to examine group-level change. Wilcoxon signed-rank tests were used to evaluate mean rank change on the items, scales, and summary risk judgments of both instruments. This test provides a nonparametric alternative to the paired-sample t-test which can be used when the data evidence non-normal distributions (Privitera, 2018; Wilcoxon, 1945). As well, Spearman's rho correlations, a nonparametric correlation coefficient examining the stability of ranked data, was used to examine relative change in ratings across assessments (Privitera, 2018).

Based on the current sample sizes across each of the assessments (i.e., *N* = 161 with at least one follow-up to *N* = 100 with all five follow-ups complete; see Table 3-1 in the Sample Size and Attrition Section for full details), the Wilcoxon signed-rank tests and Spearman's coefficients had adequate power to detect medium effects. With respect to the Wilcoxon tests, depending on the method of calculation and distribution of the variables, a sample size of between 41 to 50 is required to achieve 80% power to detect a medium effect (0.4), while a sample size between 163 to 175 is required to achieve 80% power to detect a small effect (0.2; Shieh et al., 2007). With respect to Spearman's correlations, a sample size of 85 is necessary to achieve 80% power to detect an alternative value of 0.3, while a sample size of 194 is required to achieve 80% power to detect an alternative value of 0.2 (Bujang & Baharum, 2016; May & Looney, 2020).

2.4.5. Characteristics of internal responsiveness.

To identify patterns of change, repeated measures correlations were employed. That is, repeated measures correlation was used to identify relationships amongst the scores on different dynamic ratings over time (i.e., the extent to which change on a dynamic rating is related to change on other dynamic ratings). Repeated measures correlation tests for the within-individual association of scores assessed at two or more occasions (Bakdash & Marusich, 2017; Bland & Altman, 1995a,b). This technique is an extension and atypical application of analysis of covariance (ANCOVA) that shares many characteristics with the classic Pearson correlation coefficient (e.g., bounded between -1 and 1; represents the strength of a bivariate linear relationship). Unlike other common correlations, repeated measures correlation is not based on the assumption of independence of observations, allowing for the examination of the relationship between scores from repeated assessments within participants.

The power of this technique is dependent upon sample size and the number of repeated observations (Bakdash & Marusich, 2017). Based on the current sample sizes across six assessments, these tests had more than adequate power to detect medium effects. For instance, with five repeated observations, a sample size of approximately 20 is required to achieve 80% power to detect medium effects (r = 0.3), while a sample size of roughly 200 is required to achieve 80% power to detect small effects (r = 0.1).

2.4.6. External responsiveness.

Several approaches were also employed to examine the external responsiveness of these measures (i.e., the relationship between change on the tools and change in violence). Generalized estimating equation (GEE; Diggle et al., 2002; Liang & Zeger, 1986) was employed to evaluate the extent to which change repeated assessments on the dynamic ratings was statistically associated with violence, as well as the extent to which change (scores) on the dynamic risk factors and other dynamic ratings were statistically associated with violence.

GEE is an extension of the generalized linear model that can handle non-linear functions and correlated repeated measurements within participants (Garson, 2013; Gibbons et al., 2010; Liang & Zeger, 1986). Research using repeated-measures, longitudinal designs are ideally suited for this type of analysis. It does not hold many of the (stringent) assumptions of other generalized linear modeling techniques, including the ability to examine dichotomous outcome variables. GEE is similar in many respects to generalized linear mixed modeling, although it requires no specification of the forms of any distributions and will produce consistent results under certain conditions assuming the same underlying model specifications. Thus, it presents an ideal procedure for examining the associations between change in different features of the risk assessment tools and the presence of violent behaviour.

In this context, GEE models the independent statistical association between risk ratings and violence across all six assessments reported as odds ratios for each model (Michel et al., 2013). Specifically, due to the nature of the dichotomous outcome, the current models used a binominal distribution with binary logistic link function. As well, the independent correlation structure was selected for the working correlation matrix and a robust error estimator method was used to provide superior error estimates for large datasets. These analyses included all available data, as GEE can accommodate varying numbers of observations per participant.

In this instance, the analyses examined the relationship between the dynamic risk ratings and violence reported at the next assessment (i.e., violence was lagged, so baseline ratings predict violence at the first follow-up, ratings at the first follow-up predict violence at the second follow-up, and so on). Separate models were run for each of the

dynamic risk factors, dynamic risk scales, and summary risk judgments from both the risk assessment instruments across all six assessments. To get at the incremental validity of change on these dynamic risk ratings over the baseline assessment of these ratings, additional models were run that included the corresponding baseline rating as a covariate. Finally, to address the incremental validity of these dynamic ratings over static risk ratings, additional models were run that included static risk ratings (i.e., the HCR-20 Historical scale scores) as a covariate.

Power and sample size determinations for GEE are complex and dependent upon many factors including base rates, the number of repeated observations, the strength of the measurements within participants, the correlation structure, the amount and pattern of missing data, and the variance of error terms (Jung & Ahn, 2003; Leon, 2004). Several methods have been developed for estimating sample size requirements in longitudinal studies with binary outcomes and correlated measurements (Diggle et al., 2002; Li & McKeague, 2013; Liu & Liang, 1997). Liu and Liang (1997) developed a method to determine required sample sizes for studies involving repeated correlated measurements. They present tables of sample size requirements for four repeated measurements to achieve 90% power, which roughly indicate that the current study only has sufficient power under the condition that the within-individual correlation is low (i.e., under 0.20). Additionally, Leon (2004) includes tables of sample size requirements for repeated assessments to achieve 80% power in balanced designs using the sample size calculation methods described by Diggle et al. (2002). Based on this method, with five follow-up assessments, the current sample sizes are sufficient to achieve adequate power to detect small effects (i.e., change in rate of 0.1) with an intraindividual correlation among observations of 0.20. However, with higher intraindividual correlations amongst observations (i.e., 0.4 and above), the current sample sizes are sufficient to achieve adequate power to detect medium effects (i.e., change in rate of 0.2).[10] As well, power calculations based on estimates from the current data generally indicated that the GEE analyses had adequate power to detect medium to large effects.

[10] These sample size and power estimations are based on several assumptions and estimates. Notably, they do not account for attrition and other missing data, which would result in lower power for the current analyses (Ahn, 2008).

2.4.7. Comparative responsiveness.

To evaluate the comparative responsiveness of the dynamic ratings on the two risk measures, the proportion of participants evidencing (raw score) change and reliable change on the dynamic ratings were compared separately using the test developed by McNemar (1947). This nonparametric test allows for the comparison between dependent (i.e., correlated or matched) proportions. The McNemar test is often conceptualized as a type of chi-square test for use with dependent rather than independent samples (Adedokun & Burgess, 2012). Several methods have been proposed for calculating power and sample size requirements for the McNemar test (Lachin, 1992). Based on the approach described by Duff (1984), the McNemar tests had a power of between 44% to 64%, dependent upon follow-up examined, to detect a proportion discordant of 0.1 and a discordant proportion ratio of 2.0; whereas, these tests had a power of between 74% to 91% to detect a proportion discordant of 0.2 and a discordant proportion ratio of 2.0. In addition, the stability coefficients for the various dynamic ratings were compared using the test developed by Raghunathan and colleagues (1996). This test allows for the comparison of correlated but non-overlapping correlations. Generally, these tests had adequate power (i.e., at least 80%) to detect a difference in correlation coefficients of 0.3 (Bujang & Baharum, 2016).

Chapter 3. Results

3.1. Sample Size and Attrition

The combined sample consisted of 235 participants of which 216 were assessed on the two risk assessment instruments at the baseline phase. Table 3-1 presents the number of participants that were assessed at each of the five follow-up assessments and the corresponding percent of the original sample. As seen, 74.5% of the sample was assessed at the first follow-up while 46.3% were assessed at all five follow-ups.

Also presented in Table 3-1 is the average number of days between each of the assessments and from the baseline assessment. Generally, each of the assessment was approximately one-and-a-half months apart (*M* = 38 to 52 days) and the total length of the study averaged 211 days (*SD* = 56.35). There was considerable variability in the length of each follow-up period within and across participants, as seen in the relatively large standard deviations.

Table 3-1. Attrition and Length of Re-Assessment Intervals

Assessment	Attrition		Number of Days	
	N	Percent	Per Assessment	From Baseline
Baseline	216	100%	-	-
Follow-up #1	161	74.5%	52.46 (35.05)	52.46 (35.05)
Follow-up #2	134	62.0%	46.17 (25.23)	95.95 (48.65)
Follow-up #3	119	55.1%	45.39 (28.29)	137.37 (55.13)
Follow-up #4	110	50.9%	42.91 (20.10)	175.60 (57.19)
Follow-up #5	100	46.3%	38.23 (12.25)	211.20 (56.35)

Note. Values presented in the final two columns are means and standard deviations.

Comparisons were made between participants who completed at least one follow-up and participants who did not complete any follow-ups using chi-squared tests, t-tests, and Mann-Whitney tests, as appropriate. Overall, very few differences were observed between these groups. Specifically, no differences were found across the major demographic variables (e.g., age, gender, race, marital status), current or prior

diagnoses, or number of prior psychiatric hospitalizations, as well as baseline psychopathy scores or history of violence. With respect to the HCR-20 and START, no differences were found on the scale scores, total scores, or final risk judgments. Finally, although no differences were observed for any of the items on the HCR-20, differences were seen on four START items.[11]

Comparisons were also made between participants that completed all five follow-ups (i.e., completers) and those who did not complete all follow-ups (i.e., dropouts). Once again, few differences were found between these groups. That is, no differences were found across the major demographic variables (e.g., age, gender, race, marital status), current or prior diagnoses, or number of prior psychiatric hospitalizations, as well as history of violence. However, psychopathy scores were higher in participants that did not complete all follow-ups (M = 12.76, SD = 5.82) compared to completers (M = 7.85, SD = 5.02), $t(212) = 2.58$, $p = 0.011$, $d = 0.35$, 95% CI = [0.08, 0.62]. With respect to the HCR-20 and START, no differences were found on the scale scores, total scores, or final risk judgments. However, a difference was observed in the proportions of scores for two items on the HCR-20[12] and eight items on the START[13] between these groups.

[11] Differences in the proportions of each rating (i.e., No, Possible, Present) were observed for one Strength item and three Vulnerability items. On the Strength scale, Mental State was more varied in dropouts, χ^2 (2, N = 215) = 7.33, p = .026. On the Vulnerability scale, Occupational, χ^2 (2, N = 215) = 11.64, p = .003, Substance Use, χ^2 (2, N = 217) = 6.43, p = .040, and Material Resources, χ^2 (2, N = 215) = 8.07, p = .018, were more often rated higher in dropouts.

[12] Differences in the proportions of ratings (i.e., No, Possible, Present) were observed for one Historical item and one Clinical item. Participants who dropped out of the study tended to receive higher ratings on Psychopathy, χ^2 (2, N = 216) = 11.80, p = .003, and Unresponsive to Treatment, χ^2 (2, N = 217) = 6.14, p = .047.

[13] Differences in the proportions of ratings (i.e., No, Possible, Present) were observed for seven Strength items and three Vulnerability items. On the Strength scale, Mental State was more varied in dropouts, χ^2 (2, N = 215) = 17.81, p <.001, while Social Skills, χ^2 (2, N = 216) = 8.10, p = .017, Relationships, χ^2 (2, N = 216) = 7.93, p = .019, Substance Use, χ^2 (2, N = 215) = 20.07, p <.001, Social Supports, χ^2 (2, N = 214) = 7.67, p = .022, Rule Adherence, χ^2 (2, N = 215) = 18.69, p <.001, and Treatability, χ^2 (2, N = 216) = 7.55, p = .023, were generally less present in dropouts. On the Vulnerability scale, Substance Use, χ^2 (2, N = 217) = 20.63, p <.001, Material Resources, χ^2 (2, N = 215) = 8.89, p = .012, and Rule Adherence, χ^2 (2, N = 217) = 11.12, p = .004, were rated higher in dropouts.

3.2. Descriptive Statistics across Assessments

Table 3-2 presents descriptive statistics for the HCR-20 across all assessments. Precisely, the table displays the means and standard deviations for the items, scale scores, and summary risk judgments for the baseline and each of the five follow-up assessments. The table also displays the overall values when all assessments are collapsed (N = 840). Table 3-3 presents the same information for the START.

Table 3-4 displays the percentages of the sample that perpetrated violence at each of the assessment intervals as well as any violence across any of the five follow-ups. Although not included in any of the subsequent analyses, a considerable proportion of the sample had perpetrated any violence (33.2%) in the six months prior to the baseline phase and many of these acts were more serious in nature (29.9%). As seen in the table, across each of the follow-ups approximately 11% to 19% of the sample committed at least one act of violence, whereas around 6% to 11% committed a more serious act of violence. Considering the perpetration of any violence throughout the entire study period, approximately 33% perpetrated any violence and 20% perpetrated serious violence.

Comparisons between the two subsamples on the risk assessment tools revealed several differences. Considering only the baseline assessment, the correctional sample was generally considered higher risk than the civil psychiatric sample. With respect to the HCR-20, the correctional sample had higher Historical scale scores, $t(214) = -7.59$, $p < .001$, $d = -1.07$, 95% $CI = [-1.36, -0.77]$, Risk Management scale scores, $t(214) = -7.63$, $p < .001$, $d = -1.08$, 95% $CI = [-1.37, -0.78]$, total scores, $t(214) = -7.16$, $p < .001$, $d = -1.01$, 95% $CI = [-1.30, -0.72]$, and summary risk judgments, χ^2 (2, N = 216) = 30.88, p = <.001). With respect to the START, the correctional sample had higher Vulnerability scale scores, $t(214) = 7.63$, $p < .001$, $d = -0.43$, 95% $CI = [-0.71, -0.15]$, lower Strength scale scores, $t(213) = -3.26$, $p = .001$, $d = 0.71$, 95% $CI = [0.43, 1.00]$, and higher summary risk judgments, χ^2 (2, N = 214) = 22.81, p <.001.

Table 3-2. Descriptive Statistics on the HCR-20 across Assessments

HCR-20 Rating	Baseline	Follow-up #1	Follow-up #2	Follow-up #3	Follow-up #4	Follow-up #5	Overall
Clinical (C) Scale	4.44 (2.31)	3.56 (2.30)	3.72 (2.64)	3.21 (2.14)	3.20 (2.38)	3.04 (2.35)	3.65 (2.40)
C1 Lack of Insight	0.85 (0.79)	0.75 (0.73)	0.78 (0.75)	0.66 (0.66)	0.68 (0.73)	0.65 (0.72)	0.74 (0.74)
C2 Negative Attitudes	0.61 (0.75)	0.42 (0.68)	0.51 (0.74)	0.37 (0.61)	0.36 (0.66)	0.30 (0.58)	0.45 (0.69)
C3 Active Symptoms of MMI	1.21 (0.83)	1.00 (0.83)	0.94 (0.85)	0.94 (0.80)	0.87 (0.80)	0.82 (0.83)	1.00 (0.83)
C4 Impulsivity	1.00 (0.78)	0.74 (0.71)	0.75 (0.76)	0.64 (0.70)	0.60 (0.67)	0.64 (0.69)	0.77 (0.74)
C5 Unresponsive to Tx	0.76 (0.73)	0.68 (0.71)	0.72 (0.73)	0.62 (0.66)	0.68 (0.75)	0.63 (0.65)	0.69 (0.71)
Risk Management (R) Scale	5.08 (2.40)	4.47 (2.50)	4.78 (2.66)	4.17 (2.51)	4.15 (2.63)	3.85 (2.37)	4.52 (2.53)
R1 Plans Lack Feasibility	0.81 (0.76)	0.82 (0.75)	0.87 (0.80)	0.72 (0.75)	0.71 (0.73)	0.64 (0.69)	0.77 (0.75)
R2 Exposure to Destabilizers	1.08 (0.75)	0.91 (0.77)	1.02 (0.75)	0.88 (0.76)	0.86 (0.77)	0.78 (0.71)	0.95 (0.76)
R3 Lack of Personal Support	0.93 (0.80)	0.86 (0.77)	0.87 (0.80)	0.82 (0.78)	0.79 (0.77)	0.76 (0.73)	0.85 (0.78)
R4 Noncompliance w/ Tx	0.95 (0.74)	0.66 (0.70)	0.72 (0.74)	0.56 (0.66)	0.58 (0.72)	0.63 (0.69)	0.72 (0.73)
R5 Stress	1.32 (0.67)	1.21 (0.67)	1.29 (0.66)	1.20 (0.70)	1.21 (0.67)	1.05 (0.71)	1.23 (0.68)
Summary Risk Judgment	0.65 (0.74)	0.45 (0.67)	0.47 (0.72)	0.37 (0.65)	0.35 (0.66)	0.33 (0.62)	0.47 (0.70)

Note. Values provide are means and standard deviations.

Table 3-3. Descriptive Statistics on the START across Assessments

START Rating		Baseline	Follow-up #1	Follow-up #2	Follow-up #3	Follow-up #4	Follow-up #5	Overall
Vulnerability		17.85 (7.44)	15.07 (8.18)	14.91 (8.69)	13.80 (7.89)	13.79 (8.39)	13.24 (7.74)	15.19 (8.17)
1	Social Skills	0.89 (0.73)	0.81 (0.68)	0.81 (0.66)	0.73 (0.66)	0.77 (0.63)	0.71 (0.69)	0.80 (0.68)
2	Relationships	0.87 (0.75)	0.66 (0.69)	0.63 (0.68)	0.62 (0.71)	0.58 (0.67)	0.59 (0.71)	0.68 (0.71)
3	Occupational	1.15 (0.76)	0.96 (0.77)	0.90 (0.80)	0.87 (0.75)	0.89 (0.79)	0.88 (0.74)	0.97 (0.78)
4	Recreational	0.83 (0.74)	0.99 (0.72)	0.89 (0.80)	0.82 (0.74)	0.74 (0.71)	0.87 (0.72)	0.86 (0.74)
5	Self-Care	0.50 (0.63)	0.48 (0.65)	0.50 (0.69)	0.45 (0.66)	0.44 (0.63)	0.43 (0.61)	0.47 (0.65)
6	Mental State	0.97 (0.81)	0.81 (0.78)	0.74 (0.79)	0.74 (0.80)	0.72 (0.82)	0.67 (0.81)	0.80 (0.80)
7	Emotional State	1.22 (0.75)	1.07 (0.69)	1.10 (0.73)	1.11 (0.69)	1.06 (0.67)	0.97 (0.73)	1.11 (0.72)
8	Substance Use	1.03 (0.90)	0.78 (0.86)	0.74 (0.84)	0.73 (0.85)	0.61 (0.83)	0.64 (0.82)	0.79 (0.87)
9	Impulse Control	1.00 (0.76)	0.71 (0.71)	0.72 (0.76)	0.64 (0.72)	0.62 (0.69)	0.57 (0.66)	0.75 (0.74)
10	External Triggers	0.95 (0.77)	0.80 (0.76)	0.87 (0.76)	0.75 (0.75)	0.74 (0.76)	0.73 (0.68)	0.82 (0.76)
11	Social Support	1.06 (0.80)	0.94 (0.72)	0.92 (0.75)	0.91 (0.73)	0.86 (0.79)	0.94 (0.73)	0.95 (0.76)
12	Material Resources	1.00 (0.71)	0.93 (0.71)	0.90 (0.73)	0.76 (0.72)	0.82 (0.63)	0.77 (0.69)	0.89 (0.71)
13	Attitudes	0.81 (0.76)	0.61 (0.69)	0.63 (0.72)	0.55 (0.67)	0.53 (0.66)	0.46 (0.63)	0.63 (0.71)
14	Medication Adherence	0.68 (0.73)	0.47 (0.67)	0.52 (0.65)	0.40 (0.66)	0.55 (0.73)	0.48 (0.65)	0.54 (0.69)
15	Rule Adherence	0.59 (0.74)	0.45 (0.71)	0.40 (0.70)	0.39 (0.69)	0.39 (0.69)	0.31 (0.60)	0.44 (0.70)
16	Conduct	0.70 (0.79)	0.43 (0.68)	0.42 (0.64)	0.34 (0.61)	0.28 (0.60)	0.26 (0.56)	0.45 (0.69)
17	Insight	0.89 (0.76)	0.78 (0.76)	0.80 (0.76)	0.71 (0.64)	0.78 (0.75)	0.71 (0.73)	0.79 (0.74)
18	Plans	0.92 (0.77)	0.88 (0.80)	0.87 (0.81)	0.82 (0.72)	0.89 (0.77)	0.81 (0.75)	0.87 (0.77)
19	Coping	1.14 (0.68)	0.91 (0.64)	0.99 (0.61)	0.93 (0.66)	0.90 (0.59)	0.96 (0.65)	0.99 (0.65)
20	Treatability	0.72 (0.69)	0.69 (0.71)	0.67 (0.71)	0.61 (0.63)	0.68 (0.71)	0.58 (0.64)	0.67 (0.68)
Strength		19.58 (8.66)	21.89 (8.55)	22.37 (9.11)	23.71 (8.95)	23.69 (9.20)	23.84 (9.41)	22.10 (9.04)

	START Rating	Baseline	Follow-up #1	Follow-up #2	Follow-up #3	Follow-up #4	Follow-up #5	Overall
1	Social Skills	1.18 (0.66)	1.32 (0.62)	1.29 (0.64)	1.34 (0.63)	1.35 (0.64)	1.38 (0.62)	1.29 (0.64)
2	Relationships	0.95 (0.65)	1.13 (0.66)	1.09 (0.71)	1.20 (0.71)	1.19 (0.69)	1.21 (0.64)	1.11 (0.68)
3	Occupational	0.92 (0.78)	0.93 (0.78)	0.93 (0.77)	0.96 (0.76)	0.92 (0.75)	0.96 (0.75)	0.93 (0.77)
4	Recreational	0.81 (0.71)	0.82 (0.70)	0.91 (0.70)	1.03 (0.72)	1.06 (0.69)	0.95 (0.66)	0.91 (0.70)
5	Self-Care	1.29 (0.70)	1.32 (0.75)	1.35 (0.73)	1.38 (0.69)	1.42 (0.68)	1.41 (0.70)	1.35 (0.71)
6	Mental State	1.01 (0.71)	1.18 (0.67)	1.12 (0.69)	1.22 (0.70)	1.35 (0.67)	1.27 (0.71)	1.17 (0.70)
7	Emotional State	0.82 (0.66)	0.99 (0.61)	0.96 (0.62)	1.02 (0.62)	0.96 (0.65)	0.99 (0.67)	0.94 (0.64)
8	Substance Use	0.94 (0.84)	1.12 (0.81)	1.20 (0.84)	1.27 (0.77)	1.34 (0.81)	1.33 (0.82)	1.16 (0.83)
9	Impulse Control	0.88 (0.66)	0.98 (0.66)	1.03 (0.67)	1.18 (0.70)	1.17 (0.63)	1.17 (0.70)	1.04 (0.68)
10	External Triggers	0.80 (0.73)	0.89 (0.73)	0.99 (0.77)	1.08 (0.75)	1.04 (0.71)	1.13 (0.71)	0.95 (0.74)
11	Social Support	1.04 (0.74)	1.13 (0.67)	1.18 (0.74)	1.26 (0.72)	1.16 (0.71)	1.20 (0.67)	1.14 (0.72)
12	Material Resources	0.80 (0.65)	0.87 (0.65)	0.93 (0.60)	0.94 (0.63)	1.03 (0.59)	1.01 (0.62)	0.91 (0.63)
13	Attitudes	1.00 (0.74)	1.15 (0.67)	1.20 (0.71)	1.29 (0.69)	1.29 (0.69)	1.29 (0.70)	1.17 (0.71)
14	Medication Adherence	1.16 (0.73)	1.39 (0.69)	1.45 (0.69)	1.41 (0.75)	1.38 (0.73)	1.39 (0.75)	1.34 (0.72)
15	Rule Adherence	1.22 (0.73)	1.37 (0.71)	1.40 (0.74)	1.45 (0.76)	1.55 (0.69)	1.49 (0.70)	1.39 (0.73)
16	Conduct	1.15 (0.73)	1.33 (0.70)	1.40 (0.69)	1.46 (0.71)	1.47 (0.70)	1.47 (0.67)	1.35 (0.71)
17	Insight	1.03 (0.70)	1.18 (0.67)	1.16 (0.68)	1.23 (0.68)	1.21 (0.65)	1.18 (0.69)	1.15 (0.68)
18	Plans	0.98 (0.75)	0.99 (0.69)	0.95 (0.73)	1.03 (0.71)	0.98 (0.71)	1.05 (0.73)	0.99 (0.72)
19	Coping	0.73 (0.60)	0.89 (0.60)	0.89 (0.54)	0.96 (0.54)	0.96 (0.55)	1.00 (0.59)	0.88 (0.58)
20	Treatability	1.07 (0.71)	1.18 (0.72)	1.22 (0.69)	1.26 (0.75)	1.13 (0.73)	1.19 (0.69)	1.16 (0.72)
Summary Risk Judgment		0.55 (0.71)	0.42 (0.67)	0.43 (0.68)	0.35 (0.63)	0.30 (0.62)	0.33 (0.60)	0.42 (0.67)

Note. Values provide are means and standard deviations.

Table 3-4. Violence across Assessments

Assessment	Violence	
	Broad / General	Narrow / Serious
Baseline	33.2%	29.9%
Follow-up #1	16.1%	9.3%
Follow-up #2	18.7%	11.2%
Follow-up #3	10.8%	8.3%
Follow-up #4	12.7%	7.3%
Follow-up #5	11.0%	6.1%
Any Violence	32.9%	19.9%

3.3. Interrater Reliability

Interrater reliability of the HCR-20 is presented in Table 3-5. This table displays the percent agreement and intraclass correlation coefficients with confidence intervals and categorical descriptors for the items, scale scores, and summary risk judgments. For the items, ICCs ranged considerably from .144 to .712. Notably, six items fell into the lowest category according to Cicchetti & Sparrow (1981) and two items in the lowest category according to Landis & Koch (1977). Coefficients falling into the lowest categories in each respective scale are bolded. For the scale scores, the Clinical scale yielded higher reliability (ICC = .628) than the Risk Management scale (ICC = .415). The summary risk judgments produced one of the highest coefficients (ICC = .634), falling in the good and substantial category, respectively.

Interrater reliability of the START is presented in Table 3-6. Once again, this table displays the precent agreement and intraclass correlation coefficients with confidence intervals and categorical descriptors for the items, scale scores, and summary risk judgments. For the items, ICCs ranged considerably on both the Vulnerability (ICC = .157 to .860) and Strength (ICC = .054 to .676) scales. On the Vulnerability scale, only three items fell into the lowest category according to Cicchetti & Sparrow (1981) and one item in the lowest category according to Landis & Koch (1977). In contrast, on the Strength scale, ten items fell into the lowest category according to Cicchetti & Sparrow (1981) and five items in the lowest category according to Landis & Koch (1977). Coefficients falling into the lowest categories in each respective scale are bolded. For the scale scores, the Vulnerability scale yielded higher reliability (ICC = .758) than the Strength scale (ICC = .524). The summary risk judgments produced a coefficient (ICC = .643) falling in the good and substantial category, respectively.

Table 3-5. Interrater Reliability of the HCR-20

HCR-20 Rating		Percent Agreement	ICC_1	(95% CI)	Cicchetti & Sparrow, 1981	Landis & Koch, 1977
Clinical (C) Scale		22.6%	.628	(.36 - .80)	Good	Substantial
C1	Lack of Insight	67.7%	.712	(.48 - .85)	Good	Substantial
C2	Negative Attitudes	71.0%	.368	(.02 - .64)	**Poor**	Fair
C3	Active Symptoms of MMI	54.8%	.198	(-.17 - .52)	**Poor**	**Slight**
C4	Impulsivity	58.1%	.493	(.18 - .72)	Fair	Moderate
C5	Unresponsive to Tx	29.0%	.144	(-.23 - .47)	**Poor**	**Slight**
Risk Management (R) Scale		9.7%	.415	(.09 - .66)	Fair	Moderate
R1	Plans Lack Feasibility	38.7%	.301	(-.06 - .59)	**Poor**	Fair
R2	Exposure to Destabilizers	64.5%	.667	(.41 - .83)	Good	Substantial
R3	Lack of Personal Support	34.5%	.245	(-.07 - .53)	**Poor**	Fair
R4	Noncompliance w/ Tx	41.9%	.514	(.21- .73)	Fair	Moderate
R5	Stress	51.6%	.281	(-.05 - .57)	**Poor**	Fair
Summary Risk Judgment		74.2%	.634	(.37 - .81)	Good	Substantial

Note. ICC = intraclass correlation coefficient; CI = confidence interval.

Table 3-6. Interrater Reliability of the START

START Rating		Percent Agreement	ICC_1	(95% CI)	Cicchetti & Sparrow, 1981	Landis & Koch, 1977
Vulnerability		9.7%	.758	(.56 - .88)	Excellent	Substantial
1	Social Skills	41.9%	.479	(.16 - .71)	Fair	Moderate
2	Relationships	61.3%	.483	(.17 - .71)	Fair	Moderate
3	Occupational	67.7%	.609	(.33 - .79)	Good	Substantial
4	Recreational	45.2%	.338	(-.02 - .62)	**Poor**	Fair
5	Self-Care	54.8%	.464	(.14 - .70)	Fair	Moderate
6	Mental State	38.7%	.157	(-.21 - .48)	**Poor**	**Slight**
7	Emotional State	58.1%	.454	(.12 - .70)	Fair	Moderate
8	Substance Use	77.4%	.860	(.73 - .93)	Excellent	Almost Perfect
9	Impulse Control	64.5%	.693	(.46 - .84)	Good	Substantial
10	External Triggers	61.3%	.404	(.07 - .66)	Fair	Fair
11	Social Support	48.4%	.439	(.11 - .68)	Fair	Moderate
12	Material Resources	61.3%	.550	(.25 - .75)	Fair	Moderate
13	Attitudes	71.0%	.669	(.42 - .83)	Good	Substantial
14	Medication Adherence	54.8%	.462	(.14 - .70)	Fair	Moderate
15	Rule Adherence	80.6%	.612	(.34 - .79)	Good	Substantial
16	Conduct	58.1%	.417	(.10 - .67)	Fair	Moderate
17	Insight	74.2%	.755	(.55 - .87)	Excellent	Substantial
18	Plans	58.1%	.569	(.27 - .77)	Fair	Moderate
19	Coping	43.3%	.271	(-.09 - .57)	**Poor**	Fair
20	Treatability	61.3%	.529	(.22 - .74)	Fair	Moderate
Strength		3.2%	.524	(.21 - .74)	Fair	Moderate

START Rating		Percent Agreement	ICC_1	(95% CI)	Cicchetti & Sparrow, 1981	Landis & Koch, 1977
1	Social Skills	54.8%	.566	(.28 - .76)	Fair	Moderate
2	Relationships	64.5%	.600	(.32 - .79)	Good	Moderate
3	Occupational	74.2%	.676	(.43 - .83)	Good	Substantial
4	Recreational	45.2%	.164	(-.21 - .49)	**Poor**	**Slight**
5	Self-Care	45.2%	.200	(-.17 - .52)	**Poor**	**Slight**
6	Mental State	35.5%	.318	(-.01 - .59)	**Poor**	Fair
7	Emotional State	61.3%	.526	(.22 - .74)	Fair	Moderate
8	Substance Use	67.7%	.664	(.41 - .82)	Good	Substantial
9	Impulse Control	45.2%	.063	(-.31 - .41)	**Poor**	**Slight**
10	External Triggers	41.9%	.054	(-.30 - .40)	**Poor**	**Slight**
11	Social Support	45.2%	.099	(-.26 - .43)	**Poor**	**Slight**
12	Material Resources	48.4%	.320	(-.03 - .60)	**Poor**	Fair
13	Attitudes	58.1%	.609	(.32 - .79)	Good	Substantial
14	Medication Adherence	51.6%	.479	(.16 - .71)	Fair	Moderate
15	Rule Adherence	61.3%	.371	(.02 - .64)	**Poor**	Fair
16	Conduct	54.8%	.431	(.08 - .68)	Fair	Moderate
17	Insight	64.5%	.623	(.35 - .80)	Good	Substantial
18	Plans	38.7%	.321	(-.04 - .61)	**Poor**	Fair
19	Coping	58.1%	.250	(-.12 - .55)	**Poor**	Fair
20	Treatability	54.8%	.414	(.07 - .67)	Fair	Moderate
Summary Risk Judgment		80.6%	.643	(.38 - .81)	Good	Substantial

Note. ICC = intraclass correlation coefficient; CI = confidence interval.

3.4. Reliability of Measurement

This section presents the results of the reliability of measurement analyses (i.e., SEM, MDC, floor effects and ceiling effects). These results begin to address the first research question regarding the internal responsiveness of the risk assessment tools, as internal responsiveness is dependent upon reliability and measurement error. The SEM represents how repeated measurements are distributed around a person's true score on a given scale or rating. The SEM was also used to estimate the MDC, which represents the smallest change a scale or rating is capable of detecting. Additionally, floor and ceiling effects from the baseline assessments are presented, as these participants are restricted in the ability to change due to being at the minimum or maximum rating, respectively. In the context of risk assessment ratings on the HCR-20 and START, floor effects correspond with a rating of No or Absent on the risk factors and a rating of Low Risk on the summary risk judgments; on the other hand, ceiling effects correspond with a rating of Present on the risk factors and a rating of High Risk on the summary risk judgments. Individuals with these ratings can only change in one direction.

Table 3-7 displays the reliability of measurement analyses for the HCR-20. With respect to the item ratings, the SEM tended to fall around a half-point (range = 0.397 to 0.747), while the MDC was generally around two-points (range = 1.75 to 2.40). Notably, only three items had an MDC below 2.00, which is the maximum possible range on the scale (i.e., going from Absent to Present or vice versa). As such, according to these analyses, three items (i.e., lack of insight, exposure to destabilizers, and noncompliance with treatment) can reliably detect change when the rating changes across the entire range. The summary risk judgments produced some of the smallest values compared to the item ratings, with an MDC of 1.80 indicating that only a change from Low to High (or High to Low) risk corresponds with true change in risk. For the scale scores, the MDC values indicate the scales must change by nearly four-points, which corresponds to complete change on at least two items on the five item scales, or partial change on at least four of five items. Due to the three-point rating schemes for items and summary risk judgments, floor and ceiling effects were common. Floor effects ranged from 24.5% to 54.6% for individual items and ceiling effects ranged from 15.7% to 47.0%. Most of the sample was rated Low Risk (floor effect = 50.5%), while 15.7% was rated High Risk (ceiling effect). Floor and ceiling effects were much less common on the scale scores.

Table 3-8 displays the reliability of measurement analyses for the START. With respect to the item ratings, the SEM tended to fall around a half-point (range = 0.324 to 0.738), while the MDC was generally around two-points (range = 1.58 to 2.35). Notably, 13 Vulnerability items and eight Strength items had an MDC below 2.00, which is the maximum possible range on the scale (i.e., Absent to Present or vice versa). As such, according to these analyses, approximately half the items on the START can reliably detect change when the rating changes across the entire range. The summary risk judgments again produced some of the smallest values compared to the item ratings, with an MDC of 1.75 indicating that only a change from Low to High (or High to Low) risk corresponds with true change in risk. For the scale scores, the MDC values indicate the Vulnerability scale must change by roughly four-points, which corresponds to complete change on at least two items on the twenty item scales, or partial change on at least four of twenty items; whereas, the MDC values indicate the Strength scale must change by roughly six-points, which corresponds to complete change on at least three items on the twenty item scales, or partial change on at least six of twenty items. Floor and ceiling effects were also common on the START. Floor effects ranged from 14.4% to 57.9% for individual items and ceiling effects ranged from 7.4% to 43.5%. Most of the sample was rated Low Risk (floor effect = 57.5%), while 12.6% was rated High Risk (ceiling effect). Floor and ceiling effects were much less common on the scale scores.

Table 3-7. Reliability of Measurement on the HCR-20

HCR-20 Rating		Range	SEM	MDC	Floor Effects[a]	Ceiling Effects[a]
Clinical (C) Scale		0 – 10	1.465	3.36	9 (4.2)	5 (2.3)
C1	Lack of Insight	0, 1, 2	0.397	1.75	60 (27.6)	53 (24.5)
C2	Negative Attitudes	0, 1, 2	0.549	2.05	118 (54.6)	34 (15.7)
C3	Active Symptoms of MMI	0, 1, 2	0.747	2.40	56 (25.8)	102 (47.0)
C4	Impulsivity	0, 1, 2	0.528	2.01	65 (30.0)	66 (30.4)
C5	Unresponsive to Tx	0, 1, 2	0.658	2.25	90 (41.5)	38 (17.5)
Risk Management (R) Scale		0 – 10	1.937	3.86	1 (0.5)	8 (3.7)
R1	Plans Lack Feasibility	0, 1, 2	0.629	2.20	87 (40.3)	45 (20.8)
R2	Exposure to Destabilizers	0, 1, 2	0.437	1.83	53 (24.5)	70 (32.4)
R3	Lack of Personal Support	0, 1, 2	0.676	2.28	78 (36.1)	62 (28.7)
R4	Noncompliance w/ Tx	0, 1, 2	0.505	1.97	64 (29.6)	53 (24.5)
R5	Stress	0, 1, 2	0.573	2.10	24 (11.1)	94 (43.5)
Summary Risk Judgment		0, 1, 2	0.420	1.80	109 (50.5)	34 (15.7)

Note. SEM = standard error of measurement; MDC = minimally detectable change.
[a] Values presented are sample sizes and valid percentages from the baseline assessments (*N* = 216).

Table 3-8. Reliability of Measurement on the START

START Rating		Range	SEM	MDC	Floor Effects[a]	Ceiling Effects[a]
Vulnerability		0 – 40	4.017	5.56	0 (0.0)	1 (0.5)
1	Social Skills	0, 1, 2	0.493	1.95	70 (32.4)	47 (21.8)
2	Relationships	0, 1, 2	0.513	1.99	77 (35.6)	48 (22.2)
3	Occupational	0, 1, 2	0.485	1.93	48 (22.3)	80 (37.2)
4	Recreational	0, 1, 2	0.604	2.15	79 (37.1)	43 (20.2)
5	Self-Care	0, 1, 2	0.473	1.91	125 (57.9)	16 (7.4)
6	Mental State	0, 1, 2	0.738	2.38	73 (34.0)	67 (31.2)
7	Emotional State	0, 1, 2	0.530	2.02	41 (19.1)	89 (41.4)
8	Substance Use	0, 1, 2	0.324	1.58	84 (38.7)	90 (41.5)
9	Impulse Control	0, 1, 2	0.409	1.77	61 (28.4)	61 (28.4)
10	External Triggers	0, 1, 2	0.583	2.12	69 (32.1)	58 (27.0)
11	Social Support	0, 1, 2	0.566	2.09	62 (28.7)	76 (35.2)
12	Material Resources	0, 1, 2	0.474	1.91	54 (25.1)	55 (25.6)
13	Attitudes	0, 1, 2	0.408	1.77	85 (39.9)	45 (21.1)
14	Medication Adherence	0, 1, 2	0.507	1.97	88 (47.8)	29 (15.8)
15	Rule Adherence	0, 1, 2	0.437	1.83	122 (56.2)	33 (15.2)
16	Conduct	0, 1, 2	0.527	2.01	110 (50.9)	45 (20.8)
17	Insight	0, 1, 2	0.366	1.68	74 (34.4)	51 (23.7)
18	Plans	0, 1, 2	0.508	1.98	74 (34.3)	56 (25.9)
19	Coping	0, 1, 2	0.554	2.06	37 (17.1)	67 (31.0)
20	Treatability	0, 1, 2	0.469	1.90	90 (41.7)	29 (13.4)
Strength		0 – 40	6.238	6.92	0 (0.0)	0 (0.0)

START Rating		Range	SEM	MDC	Floor Effects[a]	Ceiling Effects[a]
1	Social Skills	0, 1, 2	0.420	1.80	31 (14.4)	69 (31.9)
2	Relationships	0, 1, 2	0.428	1.81	50 (23.1)	40 (18.5)
3	Occupational	0, 1, 2	0.436	1.83	75 (34.9)	57 (26.5)
4	Recreational	0, 1, 2	0.642	2.22	77 (36.2)	37 (17.4)
5	Self-Care	0, 1, 2	0.635	2.21	31 (14.4)	94 (43.5)
6	Mental State	0, 1, 2	0.579	2.11	52 (24.2)	55 (25.6)
7	Emotional State	0, 1, 2	0.443	1.84	70 (32.6)	31 (14.4)
8	Substance Use	0, 1, 2	0.481	1.92	82 (38.3)	69 (32.2)
9	Impulse Control	0, 1, 2	0.657	2.25	62 (28.7)	36 (16.7)
10	External Triggers	0, 1, 2	0.720	2.35	83 (38.6)	39 (18.1)
11	Social Support	0, 1, 2	0.680	2.29	55 (25.7)	63 (29.4)
12	Material Resources	0, 1, 2	0.521	2.00	72 (33.5)	28 (13.0)
13	Attitudes	0, 1, 2	0.445	1.85	58 (27.0)	58 (27.0)
14	Medication Adherence	0, 1, 2	0.523	2.00	35 (19.2)	65 (35.7)
15	Rule Adherence	0, 1, 2	0.577	2.11	39 (18.0)	86 (39.6)
16	Conduct	0, 1, 2	0.539	2.03	43 (20.0)	76 (35.3)
17	Insight	0, 1, 2	0.420	1.80	50 (23.3)	56 (26.0)
18	Plans	0, 1, 2	0.593	2.13	63 (29.2)	58 (26.9)
19	Coping	0, 1, 2	0.503	1.97	76 (35.3)	17 (7.9)
20	Treatability	0, 1, 2	0.549	2.05	47 (21.8)	62 (28.7)
Summary Risk Judgment		0, 1, 2	0.398	1.75	123 (57.5)	27 (12.6)

Note. SEM = standard error of measurement; MDC = minimally detectable change.
[a] Values presented are sample sizes and valid percentages from the baseline assessments (*N* = 216).

3.5. Internal Responsiveness

This section presents the results of the main analyses addressing the first research question regarding the internal responsiveness of dynamic risk ratings. A dynamic risk factor or rating must be capable of intraindividual change over time. Three general approaches were used to examine change on the risk assessment tools. The first analyses focused on the proportions of raw score change across assessments and the number of times participants changed on the ratings across assessments.

Table 3-9 presents the frequency of change across assessments on the HCR-20. Specifically, this table displays the percent of participants that displayed no change, some change (i.e., one-point change in any direction, corresponding to change from Absent to Possible, or Possible to Present, or the opposites), or full change (i.e., two-point change in any direction, corresponding to change from Absent to Present or vice versa) across all five follow-up assessments. The table also displays the results for all pairs of assessments (N = 622), when the data were collapsed so each assessment pair (i.e., baseline to follow-up one, follow-up one to follow-up two, etc.) was included as a separate case.

As seen in Table 3-9, most ratings do not change from one assessment to the next; that is, 53% to 83% of ratings for the risk factors remained the same over time. Nevertheless, one-point change was common, occurring in 17% to 45% of participants dynamic risk factor ratings from one assessment to the next. Two-point change, on the other hand, was relatively rare occurring in 0% to 8% of item ratings. The same pattern was also evident for the summary risk judgments. Most participants maintained the same rating from one assessment to the next (71% to 87%), but one-point change (i.e., Low to Moderate, or Moderate to High, or the opposites) was common (13% to 27%). Two-point change (i.e., Low to High or vice versa) was relatively rare (0% to 2%). With respect to the scale scores, around 22% to 26% of participants maintained the same score across assessments. One-point change (33% to 36%) and two-point change (14% to 27%) were common. Change by more than two-points (on scales totalling ten with five items) was not as common occurring in 5% to 22% of participants across assessments.

Table 3-9. Participants Showing No, Some, or Full Change on the HCR-20

HCR-20 Rating		BL to FU1	FU1 to FU2	FU2 to FU3	FU3 to FU4	FU4 to FU5	Overall
Clinical (C) Scale		24, 33, 22, 21	26, 33, 25, 16	22, 42, 20, 16	26, 40, 27, 6	33, 42, 20, 5	26, 38, 23, 14
C1	Lack of Insight	62, 37, 1	68, 31, 2	60, 38, 2	66, 34, 0	69, 31, 0	65, 34, 1
C2	Negative Attitudes	65, 32, 3	78, 19, 4	74, 25, 1	78, 21, 1	83, 17, 0	75, 23, 2
C3	Active Symptoms of MMI	64, 31, 6	68, 28, 5	67, 31, 3	70, 29, 1	78, 19, 3	69, 28, 4
C4	Impulsivity	64, 34, 3	70, 27, 3	62, 36, 2	70, 28, 2	69, 29, 2	67, 31, 2
C5	Unresponsive to Tx	60, 36, 4	63, 29, 8	65, 31, 4	70, 27, 3	66, 30, 4	65, 31, 5
Risk Management (R) Scale		22, 33, 23, 22	27, 34, 22, 17	27, 44, 18, 11	25, 46, 14, 16	36, 33, 24, 7	27, 38, 20, 16
R1	Plans Lack Feasibility	60, 35, 5	67, 31, 2	69, 27, 4	66, 32, 2	69, 31, 0	66, 32, 3
R2	Exposure to Destabilizers	57, 37, 6	60, 37, 2	64, 34, 2	64, 34, 3	69, 29, 2	62, 35, 3
R3	Lack of Personal Support	67, 33, 1	62, 34, 4	73, 27, 0	82, 18, 0	82, 18, 0	72, 27, 1
R4	Noncompliance w/ Tx	53, 45, 2	65, 34, 2	73, 26, 1	72, 26, 2	69, 31, 0	65, 33, 1
R5	Stress	60, 37, 4	62, 35, 3	61, 36, 3	55, 44, 2	61, 39, 0	60, 38, 3
Summary Risk Judgment		71, 27, 2	75, 23, 2	81, 20, 0	87, 13, 1	84, 16, 0	79, 20, 1

Note. Values presented are valid percentages of no, one-point, and two-point change in either direction, except for the scale scores which show the percentages of no, one-point, two-point, and more than two-point change in either direction (values may not add to 100% due to rounding).
BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.

Table 3-10. Participants Showing No, Some, or Full Change on the START

START Rating		BL to FU1	FU1 to FU2	FU2 to FU3	FU3 to FU4	FU4 to FU5	Overall
Vulnerability		11, 13, 9, 67	11, 18, 18, 53	7, 24, 16, 53	11, 18, 17, 54	12, 15, 15, 58	11, 18, 15, 57
1	Social Skills	58, 37, 4	68, 30, 2	69, 31, 1	77, 23, 0	69, 31, 0	67, 31, 2
2	Relationships	64, 32, 4	72, 26, 2	67, 32, 1	66, 30, 4	61, 34, 5	66, 31, 3
3	Occupational	57, 37, 6	62, 37, 2	69, 30, 2	63, 36, 2	70, 29, 1	63, 34, 3
4	Recreational	57, 40, 3	56, 38, 6	64, 29, 7	64, 29, 6	60, 35, 5	60, 35, 5
5	Self-Care	69, 30, 1	78, 20, 2	75, 22, 3	69, 31, 1	71, 29, 0	72, 26, 2
6	Mental State	70, 27, 3	66, 33, 2	71, 27, 1	70, 28, 3	74, 25, 1	70, 28, 2
7	Emotional State	61, 46, 3	66, 32, 2	56, 42, 2	63, 35, 3	58, 41, 1	58, 40, 2
8	Substance Use	65, 27, 9	74, 23, 4	73, 25, 3	75, 23, 2	81, 16, 3	73, 23, 4
9	Impulse Control	59, 36, 5	62, 34, 5	63, 33, 4	62, 34, 4	64, 36, 0	62, 35, 4
10	External Triggers	60, 36, 4	56, 43, 2	61, 36, 3	57, 38, 6	57, 40, 3	58, 39, 3
11	Social Support	59, 38, 3	66, 33, 2	65, 33, 2	63, 35, 2	69, 31, 1	64, 34, 2
12	Material Resources	64, 34, 3	63, 37, 1	65, 33, 2	66, 33, 1	72, 28, 0	66, 33, 1
13	Attitudes	61, 36, 3	64, 36, 1	77, 23, 0	67, 33, 1	73, 27, 0	67, 32, 1
14	Medication Adherence	56, 38, 7	65, 31, 5	65, 32, 3	59, 33, 8	65, 34, 1	61, 34, 5
15	Rule Adherence	69, 27, 4	74, 22, 5	78, 20, 2	81, 17, 3	83, 17, 0	76, 21, 3
16	Conduct	66, 26, 8	77, 21, 2	76, 23, 1	79, 20, 1	81, 18, 1	75, 22, 3
17	Insight	66, 32, 3	62, 35, 3	64, 34, 3	64, 35, 1	70, 29, 1	65, 33, 2
18	Plans	58, 38, 5	61, 38, 2	62, 35, 3	68, 27, 6	65, 36, 0	62, 35, 3
19	Coping	49, 47, 3	59, 40, 2	64, 36, 1	58, 39, 3	67, 31, 2	59, 40, 2
20	Treatability	60, 38, 2	56, 41, 3	60, 37, 3	62, 39, 0	70, 28, 2	61, 37, 2
Strength		9, 17, 15, 59	9, 17, 19, 55	11, 19, 14, 56	10, 14, 17, 59	8, 21, 17, 54	9, 18, 16, 57
1	Social Skills	60, 39, 1	72, 28, 1	73, 27, 0	71, 28, 1	79, 20, 1	70, 30, 1

	START Rating	BL to FU1	FU1 to FU2	FU2 to FU3	FU3 to FU4	FU4 to FU5	Overall
2	Relationships	66, 33, 1	66, 33, 1	62, 37, 1	72, 27, 2	70, 28, 2	67, 32, 1
3	Occupational	57, 39, 4	63, 34, 4	64, 33, 3	64, 35, 1	68, 32, 0	63, 35, 2
4	Recreational	57, 40, 3	69, 28, 4	56, 43, 1	61, 37, 3	67, 32, 1	62, 36, 2
5	Self-Care	65, 34, 2	75, 22, 2	71, 29, 0	71, 29, 0	67, 20, 3	70, 29, 1
6	Mental State	68, 30, 2	56, 38, 5	58, 41, 2	71, 28, 1	73, 25, 2	65, 33, 2
7	Emotional State	57, 42, 1	63, 36, 2	65, 34, 1	60, 39, 2	68, 29, 3	62, 36, 2
8	Substance Use	61, 34, 6	62, 35, 3	64, 33, 3	73, 27, 1	72, 24, 4	66, 31, 4
9	Impulse Control	58, 41, 1	61, 38, 1	57, 40, 3	66, 32, 2	65, 33, 2	61, 37, 2
10	External Triggers	58, 40, 2	60, 38, 2	59, 37, 4	63, 33, 4	62, 36, 2	60, 37, 3
11	Social Support	60, 40, 0	59, 41, 0	69, 31, 1	77, 23, 0	81, 19, 0	68, 32, 0
12	Material Resources	66, 33, 1	68, 30, 2	79, 21, 0	72, 28, 0	68, 31, 1	70, 29, 1
13	Attitudes	62, 38, 1	68, 31, 1	62, 38, 1	60, 39, 1	65, 34, 1	63, 36, 1
14	Medication Adherence	55, 41, 4	64, 34, 3	70, 26, 4	67, 32, 1	65, 35, 0	63, 34, 3
15	Rule Adherence	62, 34, 4	60, 37, 3	71, 24, 5	80, 16, 5	76, 23, 1	69, 28, 4
16	Conduct	59, 38, 3	63, 37, 0	65, 34, 1	79, 16, 6	68, 32, 1	66, 32, 2
17	Insight	68, 33, 0	63, 37, 1	70, 29, 1	70, 30, 0	69, 31, 0	68, 32, 0
18	Plans	61, 37, 2	65, 35, 0	71, 25, 4	62, 36, 3	71, 29, 0	65, 33, 2
19	Coping	62, 38, 0	62, 38, 0	70, 30, 0	69, 30, 1	76, 24, 1	67, 33, 0
20	Treatability	61, 36, 3	62, 39, 0	63, 36, 2	64, 33, 3	68, 32, 0	63, 35, 2
Summary Risk Judgment		77, 20, 3	76, 22, 2	80, 19, 1	87, 12, 1	85, 15, 0	80, 18, 2

Note. Values presented are valid percentages of no, one-point, and two-point change in either direction, except for the scale scores which show the percentages of no, one-point, two-point, and more than two-point change in either direction (values may not add to 100% due to rounding).
BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.

Table 3-10 presents the same frequency of change across assessments results for the START. As seen, most ratings do not change from one assessment to the next. Generally, 49% to 83% of Vulnerability and 55% to 81% of Strength factors remained the same over time. However, one-point change was not uncommon, occurring in 17% to 47% of Vulnerability and 16% to 42% of Strength risk factor ratings across assessments. Two-point change, on the other hand, was relatively rare occurring in 0% to 9% of Vulnerability and 0% to 6% of Strength item ratings. The same pattern was also evident for the summary risk judgments. Most participants maintained the same rating from one assessment to the next (76% to 87%), but one-point change (i.e., Low to Moderate, or Moderate to High, or the opposites) was common (12% to 22%). Two-point change (i.e., Low to High or vice versa) was relatively rare (0% to 3%). With respect to the scale scores, around 7% to 12% of participants maintained the same score across assessments. One-point change (13% to 24%) and two-point change (9% to 19%) were common. Change by more than two-points (on scales totalling 40 with 20 items) was most common occurring in 53% to 67% of participants from one assessment to the next.

Table 3-11 displays the number of times ratings changed from one assessment to the next on the HCR-20. This table displays the sample size and percent of participants that did not change, changed once, twice, three times, four times, or all five times[14] across assessments for each of the ratings on the HCR-20. For this table change was considered dichotomously (i.e., no change or any change on a rating). Note that a greater number of participants had the opportunity to change once compared to more than once, and so on, based on the attrition (see Table 3-1).

It was more common for participants to change at least once on each of the ratings, except for the summary risk judgments. For the risk factors, between 23% to 49% of participants did not change across any of the assessments, meaning between 51% to 77% changed on the given item at least once across all assessments. It was common for item ratings to change once (24% to 37%) or twice (11% to 29%) across the assessments. Although it was less common for item ratings to change four (3% to 8%) or five (1% to 2%) times, some participants' item ratings did change across nearly every or every assessment. With respect to the summary risk judgments, most participants

[14] With six assessments total (i.e., baseline and five follow-up assessments), there was a maximum of five opportunities to change from one assessment to the next.

(55%) did not change across the assessments, but it was also common to change once (23%) or twice (11%) across the assessments. Again, it was less common to see participants change four (2.5%) or five (1 participant) times, but changes of this sort were observed. Due to the nature of the composite scale scores, it was much less common to see no change on the scale scores (5% and 2%). It was much more common for participants' scale scores to change at least once or even all five times (16% and 13%).

Table 3-12 displays the same results regarding the number of times ratings changed from one assessment to the next for the START. Once again, it was more common for participants to change at least once on each of the ratings across assessments, except for the summary risk judgments. For the item ratings, between 24% to 49% of participants did not change across any of the assessments, meaning between 51% to 76% changed on the given item at least once across all assessments. It was common for items to change once (21% to 34%) or twice (16% to 30%) across assessments. Some participants even changed four (2% to 11%) or five (0% to 3%) times across the study. With respect to the summary risk judgments, most participants (58%) did not change across the assessments, but it was also common to change once (20%) or twice (12%) across all assessments. Again, it was less common to see participants change four (0.6%) or five (0.6%) times. Due to the nature of the composite scale scores with 20 items on each scale, it was much less common to see no change on the scale scores (3% and 1%), and more common for participants scale scores to change four (22% and 25%) or five times (37% for both scales).

Table 3-11. Number of Times HCR-20 Ratings Changed across Assessments

HCR-20 Rating		Number of Times Changed					
		None	One	Two	Three	Four	Five
Clinical (C) Scale		8 (5.0)	28 (17.4)	29 (18.0)	34 (21.1)	37 (23.0)	25 (15.5)
C1	Lack of Insight	53 (32.9)	41 (25.5)	38 (23.6)	15 (9.3)	13 (8.1)	1 (0.6)
C2	Negative Attitudes	79 (49.1)	39 (24.2)	18 (11.2)	18 (11.2)	6 (3.7)	1 (0.6)
C3	Active Symptoms of MMI	57 (35.6)	45 (28.1)	34 (21.3)	15 (9.4)	8 (5.0)	1 (0.6)
C4	Impulsivity	53 (32.9)	44 (27.3)	36 (22.4)	21 (13.0)	6 (3.7)	1 (0.6)
C5	Unresponsive to Tx	41 (25.5)	60 (37.3)	32 (19.9)	18 (11.2)	7 (4.3)	3 (1.9)
Risk Management (R) Scale		3 (1.9)	31 (19.4)	32 (20.0)	36 (22.5)	37 (23.1)	21 (13.1)
R1	Plans Lack Feasibility	51 (31.7)	45 (28.0)	38 (23.6)	17 (10.6)	8 (5.0)	2 (1.2)
R2	Exposure to Destabilizers	47 (29.2)	45 (28.0)	32 (19.9)	24 (14.9)	11 (6.8)	2 (1.2)
R3	Lack of Personal Support	68 (42.2)	41 (25.5)	27 (16.8)	20 (12.4)	6 (3.1)	-
R4	Noncompliance w/ Tx	43 (26.7)	58 (36.0)	33 (20.5)	17 (10.6)	9 (5.6)	1 (0.6)
R5	Stress	37 (23.0)	47 (29.2)	46 (28.6)	15 (9.3)	13 (8.1)	3 (1.9)
Summary Risk Judgment		88 (54.7)	37 (23.0)	17 (10.6)	14 (8.7)	4 (2.5)	1 (0.6)

Note. Values presented are sample sizes and valid percentages. *N* = 161 participants with at least one follow-up complete.

Table 3-12. Number of Times START Ratings Changed across Assessments

START Rating		Number of Times Changed					
		None	One	Two	Three	Four	Five
Vulnerability		4 (2.5)	25 (15.5)	16 (9.9)	22 (13.7)	35 (21.7)	59 (36.6)
1	Social Skills	62 (38.5)	35 (21.7)	33 (20.5)	24 (14.9)	6 (3.7)	1 (0.6)
2	Relationships	54 (33.5)	48 (29.8)	25 (15.5)	25 (15.5)	9 (5.6)	-
3	Occupational	49 (30.4)	40 (24.8)	42 (26.1)	19 (11.8)	9 (5.6)	2 (1.2)
4	Recreational	45 (28.0)	42 (26.1)	30 (18.6)	32 (19.9)	9 (5.6)	3 (1.9)
5	Self-Care	73 (45.3)	39 (24.2)	26 (16.1)	14 (8.7)	6 (3.7)	3 (1.9)
6	Mental State	62 (38.8)	39 (24.4)	37 (23.1)	15 (9.4)	7 (4.4)	-
7	Emotional State	38 (23.8)	43 (26.9)	36 (22.5)	29 (18.1)	13 (8.1)	1 (0.6)
8	Substance Use	78 (48.4)	34 (21.1)	25 (15.5)	13 (8.1)	8 (5.0)	3 (1.9)
9	Impulse Control	55 (34.2)	34 (21.1)	33 (20.5)	24 (14.9)	10 (6.2)	5 (3.1)
10	External Triggers	41 (25.5)	37 (23.0)	46 (28.6)	20 (12.4)	15 (9.3)	2 (1.2)
11	Social Support	50 (31.1)	39 (24.2)	44 (27.3)	19 (11.8)	6 (3.7)	3 (1.9)
12	Material Resources	49 (30.4)	42 (26.1)	45 (28.0)	20 (12.4)	4 (2.5)	1 (0.6)
13	Attitudes	53 (32.9)	54 (33.5)	25 (15.5)	18 (11.2)	11 (6.8)	-
14	Medication Adherence	40 (30.3)	36 (27.3)	25 (18.9)	20 (15.2)	10 (7.6)	1 (0.8)
15	Rule Adherence	79 (49.1)	40 (24.8)	27 (16.8)	6 (3.7)	8 (5.0)	1 (0.6)
16	Conduct	78 (48.4)	41 (25.5)	22 (13.7)	12 (7.5)	6 (3.7)	2 (1.2)
17	Insight	55 (34.2)	41 (25.5)	34 (21.1)	17 (10.6)	12 (7.5)	2 (1.2)
18	Plans	49 (30.4)	45 (28.0)	26 (16.1)	27 (16.8)	11 (6.8)	3 (1.9)
19	Coping	41 (25.5)	46 (28.6)	33 (20.5)	21 (13.0)	17 (10.6)	3 (1.9)
20	Treatability	40 (24.8)	49 (30.4)	41 (25.5)	17 (10.6)	10 (6.2)	4 (2.5)
Strength		1 (0.6)	27 (16.8)	19 (11.8)	14 (8.7)	40 (24.8)	60 (37.3)
1	Social Skills	59 (36.6)	44 (27.3)	40 (24.8)	10 (6.2)	6 (3.7)	2 (1.2)

START Rating		Number of Times Changed					
		None	One	Two	Three	Four	Five
2	Relationships	57 (35.4)	33 (20.5)	48 (29.8)	16 (9.9)	6 (3.7)	1 (0.6)
3	Occupational	52 (32.3)	42 (26.1)	27 (16.8)	26 (16.1)	13 (8.1)	1 (0.6)
4	Recreational	44 (27.3)	47 (29.3)	36 (22.4)	22 (13.7)	7 (4.3)	5 (3.1)
5	Self-Care	64 (39.8)	42 (26.1)	28 (17.4)	18 (11.2)	8 (5.0)	1 (0.6)
6	Mental State	50 (31.3)	47 (29.4)	30 (18.8)	22 (13.8)	10 (6.3)	1 (0.6)
7	Emotional State	47 (29.4)	39 (24.4)	37 (23.1)	27 (16.9)	9 (5.6)	1 (0.6)
8	Substance Use	60 (37.3)	37 (23.0)	29 (18.0)	24 (14.9)	10 (6.2)	1 (0.6)
9	Impulse Control	42 (26.1)	44 (27.3)	42 (26.1)	21 (13.0)	8 (5.0)	4 (2.5)
10	External Triggers	43 (26.7)	38 (23.6)	41 (25.5)	29 (18.0)	10 (6.2)	-
11	Social Support	55 (34.2)	43 (26.7)	39 (24.2)	18 (11.2)	6 (3.7)	-
12	Material Resources	62 (38.5)	47 (29.2)	30 (18.6)	12 (7.5)	9 (5.6)	1 (0.6)
13	Attitudes	43 (26.7)	48 (29.8)	39 (24.2)	22 (13.7)	9 (5.6)	-
14	Medication Adherence	40 (30.3)	37 (28.0)	33 (25.0)	12 (9.1)	9 (6.8)	1 (0.8)
15	Rule Adherence	62 (38.5)	39 (24.2)	32 (19.9)	21 (13.0)	7 (4.3)	-
16	Conduct	51 (31.9)	45 (28.1)	35 (21.9)	23 (14.4)	3 (1.9)	3 (1.9)
17	Insight	55 (34.2)	44 (27.3)	38 (23.6)	16 (9.9)	7 (4.3)	1 (0.6)
18	Plans	53 (32.9)	44 (27.3)	33 (20.5)	21 (13.0)	8 (5.0)	2 (1.2)
19	Coping	57 (35.4)	39 (24.2)	38 (23.6)	19 (11.8)	8 (5.0)	-
20	Treatability	49 (30.4)	47 (29.2)	31 (19.3)	19 (11.8)	11 (6.8)	4 (2.5)
Summary Risk Judgment		94 (58.4)	32 (19.9)	19 (11.8)	14 (8.7)	1 (0.6)	1 (0.6)

Note. Values presented are sample sizes and valid percentages. *N* = 161 participants with at least one follow-up complete.

The following four tables (Tables 3-13 to 3-16) present the results of the RCI analyses for the HCR-20 and START. Table 3-13 presents the rate of reliable change for the HCR-20 across each of the assessment intervals, as well as the overall results for all pairs of assessments (*N* = 622). Table 3-14 presents the rate of reliable change for the HCR-20 when examining change from the baseline to each of the follow-up assessments (i.e., baseline to follow-up one, baseline to follow-up two, etc.), as opposing to examining change across each individual assessment point (i.e., baseline to follow-up one, follow-up one to follow-up two, etc.). Table 3-15 and 3-16 presents the same information for the START, respectively. All the tables display the number and percent of participants achieving reliable decreases and reliable increases.

Although change was common on both tools (as seen in Tables 3-9 to 3-12), much of this change did not reach the level of reliable change in these analyses (as seen in Tables 3-13 to 3-16). For the individual risk factor or item ratings on the HCR-20 and START to achieve reliable change, the rating needed to change completely (i.e., change from Absent to Present or vice versa), with a few exceptions. One item on the HCR-20 Clinical scale (i.e., Active Symptoms of Major Mental Illness) was unable to detect reliable change. That is, it was impossible for complete change on the item to qualify as reliable change ($Z < 1.96$ for two-point change). The same was also true for one item on the START Vulnerability scale (i.e., Mental State). In contrast, for a single item on the START Vulnerability scale (i.e., Substance Use), all change was considered reliable change. That is, any change regardless of magnitude corresponded with reliable change. Like most item ratings, the summary risk judgments made using both the HCR-20 and START needed to change completely to be classified as reliable change (i.e., change from Low to High or vice versa). Thus, the proportions of the sample showing reliable change on these ratings ignore partial change (expect for a single item on the START Vulnerability scale), while partial change was seen to be the most common type of change in Tables 3-9 and 3-10.

Table 3-13. Reliable Change on the HCR-20 across Assessments

HCR-20 Rating		BL to FU1		FU1 to FU2		FU2 to FU3		FU3 to FU4		FU4 to FU5		Overall	
		RD	RI	RD	RI	RD	RI	RD	RI	RD	RI	RD	RI
Clinical (C) Scale		4 (2.5)	0	2 (1.5)	1 (0.7)	0	1 (0.8)	0	0	0	1 (1.0)	6 (1.0)	3 (0.5)
C1	Lack of Insight	0	2 (1.2)	2 (1.5)	0	2 (1.7)	0	0	0	0	0	4 (0.6)	2 (0.3)
C2	Negative Attitudes	4 (2.5)	0	1 (0.7)	4 (3.0)	1 (0.8)	0	1 (0.9)	0	0	0	7 (1.1)	4 (0.6)
C3	Active Symptoms of MMI	-	-	-	-	-	-	-	-	-	-	-	-
C4	Impulsivity	4 (2.5)	1 (0.6)	2 (1.5)	2 (1.5)	2 (1.7)	0	1 (0.9)	1 (0.9)	0	2 (2.0)	9 (1.4)	6 (1.0)
C5	Unresponsive to Tx	2 (1.2)	4 (2.5)	4 (3.0)	6 (4.5)	2 (1.7)	3 (2.5)	1 (0.9)	2 (1.8)	3 (3.0)	1 (1.0)	12 (1.9)	16 (2.6)
Risk Management (R) Scale		3 (1.9)	1 (0.6)	1 (0.7)	2 (1.5)	1 (0.8)	0	0	0	0	0	5 (0.8)	3 (0.5)
R1	Plans Lack Feasibility	4 (2.5)	4 (2.5)	1 (0.7)	1 (0.7)	4 (3.4)	1 (0.8)	1 (0.9)	1 (0.9)	0	0	10 (1.6)	7 (1.1)
R2	Exposure to Destabilizers	7 (4.3)	2 (1.2)	1 (0.7)	2 (1.5)	1 (0.8)	1 (0.8)	0	3 (2.7)	1 (1.0)	1 (1.0)	10 (1.6)	9 (1.4)
R3	Lack of Personal Support	1 (0.6)	0	1 (0.7)	4 (3.0)	0	0	0	0	0	0	2 (0.3)	4 (0.6)
R4	Noncompliance w/ Tx	3 (1.9)	0	1 (0.7)	1 (0.7)	1 (0.8)	0	0	2 (1.8)	0	0	5 (0.8)	3 (0.5)
R5	Stress	5 (3.1)	1 (0.6)	0	4 (3.0)	3 (2.5)	1 (0.8)	0	2 (1.8)	0	0	8 (1.3)	8 (1.3)
Summary Risk Judgment		2 (1.2)	2 (1.2)	1 (0.7)	1 (0.7)	0	0	0	1 (0.9)	0	0	3 (0.5)	4 (0.6)

Note. Values presented are sample size and valid percentage. RD = reliable decrease; RI = reliable increase; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.

Table 3-14. Reliable Change on the HCR-20 from Baseline

HCR-20 Rating		BL to FU1		BL to FU2		BL to FU3		BL to FU4		BL to FU5	
		RD	RI	RD	RI	RD	RI	RD	RI	RD	RI
Clinical (C) Scale		4 (2.5)	0	5 (3.7)	4 (3.0)	2 (1.7)	0	6 (5.5)	1 (0.9)	5 (5.0)	1 (1.0)
C1	Lack of Insight	0	2 (1.2)	4 (3.0)	4 (3.0)	2 (1.7)	0	3 (2.7)	3 (2.7)	5 (5.0)	2 (2.0)
C2	Negative Attitudes	4 (2.5)	0	1 (0.7)	2 (1.5)	2 (1.7)	0	4 (3.6)	1 (0.9)	3 (3.0)	0
C3	Active Symptoms of MMI	-	-	-	-	-	-	-	-	-	-
C4	Impulsivity	4 (2.5)	1 (0.6)	8 (6.0)	0	6 (5.0)	1 (0.8)	7 (6.4)	0	7 (7.0)	1 (1.0)
C5	Unresponsive to Tx	2 (1.2)	4 (2.5)	3 (2.2)	4 (3.0)	5 (4.2)	0	4 (3.6)	6 (5.5)	3 (3.0)	2 (2.0)
Risk Management (R) Scale		3 (1.9)	1 (0.6)	1 (0.7)	0	1 (0.8)	0	2 (1.8)	0	2 (2.0)	0
R1	Plans Lack Feasibility	4 (2.5)	4 (2.5)	3 (2.2)	2 (1.5)	2 (1.7)	3 (2.5)	4 (3.6)	4 (3.6)	3 (3.0)	2 (2.0)
R2	Exposure to Destabilizers	7 (4.3)	2 (1.2)	4 (3.0)	1 (0.7)	5 (4.2)	1 (0.8)	2 (1.8)	1 (0.9)	5 (5.0)	1 (1.0)
R3	Lack of Personal Support	1 (0.6)	0	3 (2.2)	1 (0.7)	2 (1.7)	0	3 (2.7)	0	4 (4.0)	0
R4	Noncompliance w/ Tx	3 (1.9)	0	4 (3.0)	1 (0.7)	4 (3.4)	1 (0.8)	7 (6.4)	1 (0.9)	4 (4.0)	0
R5	Stress	5 (3.1)	1 (0.6)	2 (1.5)	2 (1.5)	5 (4.2)	3 (2.5)	2 (1.8)	1 (0.9)	7 (7.0)	1 (1.0)
Summary Risk Judgment		2 (1.2)	2 (1.2)	4 (3.0)	1 (0.7)	3 (2.5)	0	4 (3.6)	1 (0.9)	3 (3.0)	1 (1.0)

Note. Values presented are sample size and valid percentage. RD = reliable decrease; RI = reliable increase; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.

Table 3-15. Reliable Change on the START across Assessments

START Rating		BL to FU1		FU1 to FU2		FU2 to FU3		FU3 to FU4		FU4 to FU5		Overall	
		RD	RI	RD	RI	RD	RI	RD	RI	RD	RI	RD	RI
Vulnerability		14 (8.8)	0	2 (1.5)	1 (0.7)	2 (1.7)	1 (0.8)	3 (2.8)	3 (2.8)	0	1 (1.0)	21 (3.4)	6 (1.0)
1	Social Skills	6 (3.7)	1 (0.6)	1 (0.7)	2 (1.5)	1 (0.8)	0	0	0	0	0	8 (1.3)	3 (0.5)
2	Relationships	5 (3.1)	1 (0.6)	3 (2.3)	0	1 (0.8)	0	3 (2.8)	1 (0.9)	1 (1.0)	4 (4.0)	13 (2.1)	6 (1.0)
3	Occupational	6 (3.8)	3 (1.9)	1 (0.7)	1 (0.7)	2 (1.7)	0	1 (0.9)	1 (0.9)	0	1 (1.0)	10 (1.6)	6 (1.0)
4	Recreational	0	5 (3.1)	6 (4.5)	2 (1.5)	5 (4.2)	3 (2.5)	3 (2.8)	4 (3.7)	2 (2.0)	3 (3.0)	16 (2.6)	17 (2.7)
5	Self-Care	1 (0.6)	1 (0.6)	1 (0.8)	2 (1.5)	2 (1.7)	2 (1.7)	1 (0.9)	0	0	0	5 (0.8)	5 (0.8)
6	Mental State	-	-	-	-	-	-	-	-	-	-	-	-
7	Emotional State	5 (3.1)	0	1 (0.7)	2 (1.5)		2 (1.7)	2 (1.8)	1 (0.9)	0	1 (1.0)	8 (1.3)	6 (1.0)
8	Substance Usea	38 (24)	19 (12)	18 (13)	17 (13)	12 (10)	20 (17)	19 (17)	8 (7)	6 (6)	13 (13)	93 (15)	77 (12)
9	Impulse Control	7 (4.4)	1 (0.6)	1 (0.8)	5 (3.8)	4 (3.4)	1 (0.8)	1 (0.9)	3 (2.8)	0	0	13 (2.1)	10 (1.6)
10	External Triggers	5 (3.1)	1 (0.6)	2 (1.5)	0	1 (0.8)	2 (1.7)	3 (2.8)	3 (2.8)	1 (1.0)	2 (2.0)	12 (1.9)	8 (1.3)
11	Social Support	4 (2.5)	1 (0.6)	2 (1.5)	0	0	2 (1.7)	0	2 (1.8)	1 (1.0)	0	7 (1.1)	5 (0.8)
12	Material Resources	2 (1.3)	2 (1.3)	1 (0.7)	0	1 (0.8)	1 (0.8)	1 (0.9)	0	0	0	5 (0.8)	3 (0.5)
13	Attitudes	4 (2.5)	1 (0.6)	0	1 (0.7)	0	0	1 (0.9)	0	0	0	5 (0.8)	2 (0.3)
14	Medication Adherence	7 (5.5)	2 (1.6)	2 (1.9)	3 (2.8)	2 (2.1)	1 (1.1)	1 (1.2)	6 (7.1)	0	1 (1.3)	12 (2.4)	13 (2.6)
15	Rule Adherence	5 (3.1)	2 (1.2)	3 (2.2)	3 (2.2)	0	2 (1.7)	3 (2.8)	0	0	0	11 (1.8)	7 (1.1)
16	Conduct	12 (7.5)	1 (0.6)	2 (1.5)	1 (0.7)	0	1 (0.8)	1 (0.9)	0	0	1 (1.0)	15 (2.4)	4 (0.6)
17	Insight	2 (1.3)	3 (1.9)	2 (1.5)	2 (1.5)	3 (2.5)	0	0	1 (0.9)	1 (1.0)	0	8 (1.3)	6 (1.0)
18	Plans	5 (3.1)	3 (1.9)	1 (0.7)	1 (0.7)	3 (2.5)	1 (0.8)	1 (0.9)	5 (4.6)	0	0	10 (1.6)	10 (1.6)
19	Coping	5 (3.1)	0	1 (0.7)	1 (0.7)	1 (0.8)	0	3 (2.8)	0	1 (1.0)	1 (1.0)	11 (1.8)	2 (0.3)
20	Treatability	2 (1.2)	1 (0.6)	2 (1.5)	2 (1.5)	2 (1.7)	1 (0.8)	0	0	2 (2.0)	0	8 (1.3)	4 (0.6)
Strength		1 (0.6)	0	0	1 (0.7)	0	2 (1.7)	0	0	0	0	1 (0.2)	3 (0.5)
1	Social Skills	0	2 (1.2)	0	1 (0.7)	0	0	0	1 (0.9)	0	1 (1.0)	0	5 (0.8)

	START Rating	BL to FU1		FU1 to FU2		FU2 to FU3		FU3 to FU4		FU4 to FU5		Overall	
		RD	RI	RD	RI	RD	RI	RD	RI	RD	RI	RD	RI
2	Relationships	0	2 (1.3)	0	1 (0.8)	0	1 (0.8)	1 (0.9)	1 (0.9)	1 (1.0)	1 (1.0)	2 (0.3)	6 (1.0)
3	Occupational	4 (2.5)	2 (1.3)	2 (1.5)	3 (2.3)	3 (2.5)	0	0	1 (0.9)	0	0	9 (1.5)	6 (1.0)
4	Recreational	2 (1.3)	2 (1.3)	1 (0.7)	4 (3.0)	0	1 (0.8)	1 (0.9)	2 (1.8)	1 (1.0)	0	5 (0.8)	9 (1.5)
5	Self-Care	3 (1.9)	0	1 (0.7)	2 (1.5)	0	0	0	0	2 (2.0)	1 (1.0)	6 (1.0)	3 (0.5)
6	Mental State	1 (0.6)	2 (1.3)	3 (2.3)	4 (3.0)	1 (0.8)	1 (0.8)	0	1 (0.9)	1 (1.0)	1 (1.0)	6 (1.0)	9 (1.5)
7	Emotional State	0	2 (1.3)	2 (1.5)	0	1 (0.8)	0	1 (0.9)	1 (0.9)	3 (3.0)	0	7 (1.1)	3 (0.5)
8	Substance Use	2 (1.3)	7 (4.4)	1 (0.7)	3 (2.2)	2 (1.7)	2 (1.7)	0	1 (0.9)	4 (4.1)	0	9 (1.5)	13 (2.1)
9	Impulse Control	1 (0.6)	1 (0.6)	0	1 (0.7)	0	4 (3.4)	1 (0.9)	1 (0.9)	2 (2.0)	0	4 (0.6)	7 (1.1)
10	External Triggers	1 (0.6)	2 (1.3)	0	2 (1.5)	1 (0.8)	4 (3.4)	2 (1.8)	2 (1.8)	0	2 (2.0)	4 (0.6)	12 (1.9)
11	Social Support	0	0	0	0	0	1 (0.8)	0	0	0	0	0	1 (0.2)
12	Material Resources	1 (0.6)	0	1 (0.7)	2 (1.5)	0	0	0	0	0	1 (1.0)	2 (0.3)	3 (0.5)
13	Attitudes	0	1 (0.6)	0	1 (0.7)	0	1 (0.9)	1 (0.9)	0	0	1 (1.0)	1 (0.2)	4 (0.6)
14	Medication Adherence	0	5 (4.0)	2 (1.9)	1 (0.9)	3 (3.2)	1 (1.1)	1 (1.2)	0	0	0	6 (1.2)	7 (1.4)
15	Rule Adherence	2 (1.2)	5 (3.1)	3 (2.2)	1 (0.7)	4 (3.4)	2 (1.7)	0	5 (4.6)	1 (1.0)	0	10 (1.6)	13 (2.1)
16	Conduct	1 (0.6)	4 (2.5)	0	0	1 (0.8)	0	1 (0.9)	5 (4.6)	1 (1.0)	0	4 (0.6)	9 (1.5)
17	Insight	0	0	1 (0.7)	0	0	1 (0.8)	0	0	0	0	1 (0.2)	1 (0.2)
18	Plans	2 (1.2)	1 (0.6)	0	0	1 (0.8)	4 (3.4)	3 (2.8)	0	0	0	6 (1.0)	5 (0.8)
19	Coping	0	0	0	0	0	0	1 (0.9)	0	0	1 (1.0)	1 (0.2)	1 (0.2)
20	Treatability	1 (0.6)	4 (2.5)	0	0	0	2 (1.7)	3 (2.8)	0	0	0	4 (0.6)	6 (1.0)
Summary Risk Judgment		3 (1.9)	2 (1.3)	2 (1.5)	1 (0.8)	1 (0.9)	0	0	1 (0.9)	0	0	6 (1.0)	4 (0.7)

Note. Values presented are sample size and valid percentage. RD = reliable decrease; RI = reliable increase; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.

[a] Reliable change on this item corresponded with any change in rating, whereas full change was required for all other items.

Table 3-16. Reliable Change on the START from Baseline

START Rating		BL to FU1		BL to FU2		BL to FU3		BL to FU4		BL to FU5	
		RD	RI	RD	RI	RD	RI	RD	RI	RD	RI
Vulnerability		14 (8.8)	0	13 (9.8)	0	12 (10.2)	0	11 (10.2)	1 (0.9)	10 (10.1)	0
1	Social Skills	6 (3.7)	1 (0.6)	4 (3.0)	4 (3.0)	3 (2.5)	0	3 (2.8)	0	4 (4.0)	0
2	Relationships	5 (3.1)	1 (0.6)	6 (4.5)	0	6 (5.0)	0	6 (5.5)	2 (1.8)	7 (7.0)	1 (1.0)
3	Occupational	6 (3.8)	3 (1.9)	5 (3.8)	2 (1.5)	3 (2.5)	2 (1.7)	6 (5.6)	2 (1.9)	4 (4.0)	1 (1.0)
4	Recreational	0	5 (3.1)	2 (1.5)	4 (3.0)	3 (2.6)	2 (1.7)	7 (6.5)	4 (3.7)	2 (2.0)	4 (4.1)
5	Self-Care	1 (0.6)	1 (0.6)	1 (0.7)	5 (3.7)	3 (2.5)	4 (3.4)	1 (0.9)	1 (0.9)	2 (2.0)	1 (1.0)
6	Mental State	-	-	-	-	-	-	-	-	-	-
7	Emotional State	5 (3.1)	0	6 (4.5)	3 (2.3)	4 (3.4)	1 (0.8)	4 (3.7)	4 (3.7)	5 (5.1)	1 (1.0)
8	Substance Use[a]	38 (24)	19 (12)	34 (25)	15 (11)	24 (20)	16 (13)	26 (24)	13 (12)	21 (21)	15 (15)
9	Impulse Control	7 (4.4)	1 (0.6)	5 (3.8)	1 (0.8)	8 (6.8)	0	4 (3.7)	0	5 (5.0)	0
10	External Triggers	5 (3.1)	1 (0.6)	3 (2.3)	0	4 (3.4)	1 (0.8)	4 (3.7)	1 (0.9)	2 (2.0)	1 (1.0)
11	Social Support	4 (2.5)	1 (0.6)	2 (1.5)	1 (0.7)	0	0	2 (1.8)	1 (0.9)	4 (4.0)	3 (3.0)
12	Material Resources	2 (1.3)	2 (1.3)	2 (1.5)	0	2 (1.7)	2 (1.7)	0	1 (0.9)	1 (1.0)	2 (2.1)
13	Attitudes	4 (2.5)	1 (0.6)	5 (3.8)	1 (0.8)	5 (4.3)	1 (0.9)	5 (4.6)	1 (0.9)	5 (5.1)	0
14	Medication Adherence	7 (5.5)	2 (1.6)	5 (4.6)	0	8 (8.3)	2 (2.1)	8 (9.6)	3 (3.6)	3 (3.8)	0
15	Rule Adherence	5 (3.1)	2 (1.2)	4 (3.0)	0	3 (2.5)	1 (0.8)	4 (3.7)	0	4 (4.0)	2 (2.0)
16	Conduct	12 (7.5)	1 (0.6)	12 (9.0)	0	8 (6.8)	0	12 (11.1)	0	11 (11.1)	0
17	Insight	2 (1.3)	3 (1.9)	3 (2.3)	5 (3.8)	3 (2.5)	0	2 (1.9)	2 (1.9)	3 (3.0)	1 (1.0)
18	Plans	5 (3.1)	3 (1.9)	9 (6.7)	3 (2.2)	4 (3.4)	2 (1.7)	4 (3.7)	6 (5.5)	3 (3.0)	2 (2.0)
19	Coping	5 (3.1)	0	2 (1.5)	1 (0.7)	6 (5.0)	0	6 (5.5)	1 (0.9)	2 (2.0)	1 (1.0)
20	Treatability	2 (1.2)	1 (0.6)	3 (2.2)	3 (2.2)	4 (3.4)	1 (0.8)	3 (2.8)	2 (1.8)	1 (1.0)	1 (1.0)
Strength		1 (0.6)	0	1 (0.8)	1 (0.8)	1 (0.8)	1 (0.8)	2 (1.9)	2 (1.9)	1 (1.0)	0
1	Social Skills	0	2 (1.2)	1 (0.7)	1 (0.7)	0	1 (0.8)	1 (0.9)	3 (2.8)	0	1 (1.0)

	START Rating	BL to FU1		BL to FU2		BL to FU3		BL to FU4		BL to FU5	
		RD	RI	RD	RI	RD	RI	RD	RI	RD	RI
2	Relationships	0	2 (1.3)	2 (1.5)	1 (0.7)	0	2 (1.7)	1 (0.9)	1 (0.9)	0	2 (2.0)
3	Occupational	4 (2.5)	2 (1.3)	3 (2.3)	1 (0.8)	4 (3.4)	2 (1.7)	2 (1.9)	3 (2.8)	3 (3.0)	2 (2.0)
4	Recreational	2 (1.3)	2 (1.3)	1 (0.8)	3 (2.3)	2 (1.7)	4 (3.4)	3 (2.8)	8 (7.5)	0	3 (3.1)
5	Self-Care	3 (1.9)	0	2 (1.5)	0	2 (1.7)	1 (0.8)	1 (0.9)	1 (0.9)	1 (1.0)	1 (1.0)
6	Mental State	1 (0.6)	2 (1.3)	3 (2.3)	3 (2.3)	2 (1.7)	2 (1.7)	0	3 (2.8)	1 (1.0)	4 (4.0)
7	Emotional State	0	2 (1.3)	3 (2.3)	3 (2.3)	1 (0.8)	3 (2.5)	1 (0.9)	5 (4.6)	2 (2.0)	4 (4.0)
8	Substance Use	2 (1.3)	7 (4.4)	3 (2.3)	7 (5.3)	2 (1.7)	5 (4.3)	1 (0.9)	8 (7.5)	4 (4.1)	4 (4.1)
9	Impulse Control	1 (0.6)	1 (0.6)	1 (0.7)	2 (1.5)	0	6 (5.0)	0	5 (4.6)	0	4 (4.0)
10	External Triggers	1 (0.6)	2 (1.3)	0	4 (3.0)	0	11 (9.3)	0	7 (6.5)	0	6 (6.1)
11	Social Support	0	0	1 (0.8)	2 (1.5)	0	2 (1.7)	0	0	0	2 (2.1)
12	Material Resources	1 (0.6)	0	1 (0.8)	3 (2.3)	1 (0.8)	2 (1.7)	0	1 (0.9)	1 (1.0)	1 (1.0)
13	Attitudes	0	1 (0.6)	0	2 (1.5)	0	3 (2.5)	1 (0.9)	4 (3.7)	1 (1.0)	1 (1.0)
14	Medication Adherence	0	5 (4.0)	1 (0.9)	8 (7.5)	1 (1.1)	5 (5.3)	1 (1.2)	3 (3.7)	3 (3.8)	4 (5.1)
15	Rule Adherence	2 (1.2)	5 (3.1)	1 (0.7)	6 (4.5)	3 (2.5)	5 (4.2)	1 (0.9)	5 (4.6)	2 (2.0)	3 (3.0)
16	Conduct	1 (0.6)	4 (2.5)	1 (0.8)	3 (2.3)	1 (0.8)	5 (4.2)	1 (0.9)	5 (4.7)	0	6 (6.1)
17	Insight	0	0	1 (0.8)	2 (1.5)	0	2 (1.7)	1 (0.9)	1 (0.9)	0	2 (2.0)
18	Plans	2 (1.2)	1 (0.6)	1 (0.7)	2 (1.5)	2 (1.7)	1 (0.8)	4 (3.7)	2 (1.8)	2 (2.0)	1 (1.0)
19	Coping	0	0	0	0	0	2 (1.7)	2 (1.9)	0	1 (1.0)	1 (1.0)
20	Treatability	1 (0.6)	4 (2.5)	2 (1.5)	0	1 (0.8)	3 (2.5)	2 (1.8)	1 (0.9)	0	1 (1.0)
Summary Risk Judgment		3 (1.9)	2 (1.3)	4 (3.0)	0	3 (2.6)	0	5 (4.7)	0	2 (2.0)	0

Note. Values presented are sample size and valid percentage. RD = reliable decrease; RI = reliable increase; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.

[a] Reliable change on this item corresponded with any change in rating, whereas full change was required for all other items.

Focusing on the ratings across assessment intervals (Tables 3-13 and 3-15), the largest rates of reliable change occurred between the baseline assessment and the first follow-up. More participants generally experienced reliable decreases in risk factor ratings and reliable increases in strength ratings from baseline to the first follow-up compared to the other assessment intervals. After the baseline assessment, no clear patterns appeared in the direction of change across the follow-up assessments. There also appeared be to some variability in the rates of reliable change across items on the same instrument. For instance, on the HCR-20, the items Unresponsive to Treatment and Exposure to Destabilizers yielded potentially larger rates of reliable change than other items. Also, the HCR-20 scale scores experienced relatively equivalent rates of reliable change as most items; whereas the START Vulnerability scale yielded some of the highest rates of change for this tool and the Strength scale some of the lowest. Across instruments, the largest rates of reliable change were yielded by the START Vulnerability items, followed by the HCR-20 items, and then START Strength items.

Focusing on ratings from the baseline to each follow-up (Tables 3-14 and 3-16), although the length of interval increases considerably (see Table 3-1), there was no obvious increase (or pattern) in the rates of reliable change when comparing across assessment intervals. The rates of reliable change varied slightly across the intervals. Evident from these comparisons is the frequent changes between each assessment interval experienced by some participants, as seen in the number of times participants changed across assessments in Tables 3-11 and 3-12. The same variability in rates of reliable change across items on the same instrument was apparent when considering change from the baseline assessment.

The next four tables (Tables 3-17 to 3-20) present the results of the aggregate-level analyses of change for the HCR-20 and START. Table 3-17 presents the results of the Wilcoxon signed-rank tests and stability coefficients for the HCR-20 across each of the assessment intervals, as well as the overall results for all pairs of assessments (N = 622). Table 3-18 presents the same results for the HCR-20 when examining change from the baseline assessment to each of the follow-up assessments (i.e., baseline to follow-up one, baseline to follow-up two, etc.). Table 3-19 and 3-20 presents the same results for the START, respectively. All the tables display the Wilcoxon Z-coefficients and Spearman's stability coefficients with significant values indicated.

Table 3-17. Aggregate-Level Change on the HCR-20 across Assessments

HCR-20 Rating	BL to FU1		FU1 to FU2		FU2 to FU3		FU3 to FU4		FU4 to FU5		Overall	
	Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s
Clinical (C) Scale	–5.59***	.698***	–0.36	.683***	–1.92	.694***	–0.59	.769***	–0.07	.776***	–3.53***	.713***
C1 Lack of Insight	–1.34	.647***	–0.42	.657***	–1.51	.550***	–0.49	.647***	–0.18	.637***	–0.98	.631***
C2 Negative Attitudes	–4.12***	.621***	–1.08	.691***	–2.06*	.592***	–0.19	.657***	–0.24	.686***	–2.81**	.641***
C3 Active Symptoms of MMI	–4.00***	.641***	–1.94	.684***	–0.29	.693***	–1.33	.748***	–0.90	.767***	–3.91***	.700***
C4 Impulsivity	–4.86***	.672***	–0.00	.627***	–0.99	.560***	–0.16	.583***	–1.16	.553***	–2.57*	.611***
C5 Unresponsive to Tx	–1.07	.481***	–0.58	.442***	–0.77	.515***	–1.23	.630***	–0.03	.513***	–0.14	.507***
Risk Management (R) Scale	–3.71***	.601***	–1.84	.688***	–3.12**	.778***	–0.36	.688***	–0.68	.776***	–2.74**	.696***
R1 Plans Lack Feasibility	–0.31	.521***	–0.85	.692***	–1.95	.647***	–0.15	.632***	–0.54	.681***	–0.93	.628***
R2 Exposure to Destabilizers	–3.05**	.527***	–2.28*	.612***	–2.02*	.649***	–0.40	.613***	–0.82	.637***	–1.97*	.592***
R3 Lack of Personal Support	–1.19	.714***	–0.27	.582***	–0.35	.770***	–0.45	.826***	–0.00	.836***	–0.79	.733***
R4 Noncompliance w/ Tx	–5.24***	.566***	–1.23	.597***	–2.20***	.680***	–0.50	.619***	–1.98*	.619***	–2.45*	.595***
R5 Stress	–1.45	.413***	–1.16	.467***	–1.32	.461***	–0.01	.452***	–2.08*	.595***	–1.55	.470***
Summary Risk Judgment	–4.00***	.650***	–0.48	.653***	–1.88	.744***	–0.47	.791***	–1.50	.691***	–2.64**	.692***

Note. Z = Z-coefficient from Wilcoxon signed-rank test (Z-coefficients are always negative); r_s = Spearman's rho stability coefficient; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3-18. Aggregate-Level Change on the HCR-20 from Baseline

HCR-20 Rating		BL to FU1		BL to FU2		BL to FU3		BL to FU4		BL to FU5	
		Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s
Clinical (C) Scale		–5.59***	.698***	–3.46***	.544***	–5.40***	.593***	–4.38***	.508***	–4.15***	.450***
C1	Lack of Insight	–1.34	.647***	–0.11	.497***	–1.79	.586***	–0.86	.507***	–1.31	.478***
C2	Negative Attitudes	–4.12***	.621***	–1.88	.635***	–3.78***	.511***	–2.98**	.383***	–3.25**	.387***
C3	Active Symptoms of MMI	–4.00***	.641***	–4.44***	.576***	–4.32***	.586***	–4.33***	.484***	–4.38***	.415***
C4	Impulsivity	–4.86***	.672***	–3.94***	.500***	–5.07***	.515***	–4.83***	.468***	–3.26**	.347***
C5	Unresponsive to Tx	–1.07	.481***	–0.22	.385***	–0.64	.364***	–0.57	.297**	–0.01	.227*
Risk Management (R) Scale		–3.71***	.601***	–1.08	.621***	–4.02***	.618***	–3.48***	.619***	–4.03***	.531***
R1	Plans Lack Feasibility	–0.31	.521***	–0.68	.562***	–1.02	.453***	–1.09	.382***	–1.85	.500***
R2	Exposure to Destabilizers	–3.05**	.527***	–0.92	.509***	–2.40*	.530***	–2.47*	.578***	–2.85**	.462***
R3	Lack of Personal Support	–1.19	.714***	–0.39	.635***	–1.81	.711***	–1.58	.638***	–1.51	.531***
R4	Noncompliance w/ Tx	–5.24***	.566***	–3.73***	.547***	–5.23***	.518***	–4.04***	.371***	–3.21**	.426***
R5	Stress	–1.45	.413***	–0.12	.405***	–1.45	.302***	–1.21	.298**	–3.05**	.373***
Summary Risk Judgment		–4.00***	.650***	–3.14**	.604***	–5.09***	.700***	–4.58***	.628***	–4.01***	.568***

Note. Z = Z-coefficient from Wilcoxon signed-rank test (Z-coefficients are always negative); r_s = Spearman's rho stability coefficient; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3-19. Aggregate-Level Change on the START across Assessments

START Rating		BL to FU1		FU1 to FU2		FU2 to FU3		FU3 to FU4		FU4 to FU5		Overall	
		Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s
Vulnerability		–5.32***	.668***	–0.03	.794***	–1.45	.782***	–0.27	.700***	–0.13	.775***	–3.49***	.737***
1	Social Skills	–1.94	.456***	–0.56	.578***	–0.95	.617***	–1.00	.717***	–0.73	.641***	–1.25	.581***
2	Relationships	–4.02***	.600***	–1.20	.646***	0.00	.590***	–0.45	.499***	–0.85	.400***	–2.46*	.559***
3	Occupational	–1.54	.486***	–0.66	.649***	–0.47	.694***	–0.44	.622***	–0.17	.717***	–1.27	.624***
4	Recreational	–2.42*	.529***	–1.45	.485***	–1.22	.524***	–0.72	.479***	–1.47	.436***	–0.22	.493***
5	Self-Care	–0.27	.524***	–1.44	.637***	–0.61	.592***	–0.16	.492***	–0.38	.549***	0.00	.560***
6	Mental State	–3.52***	.728***	–1.94	.678***	0.00	.711***	–0.34	.702***	–0.93	.736***	–3.24**	.713***
7	Emotional State	–3.22**	.458***	–0.41	.595***	–1.05	.517***	–0.57	.504***	–1.94	.553***	–2.13*	.515***
8	Substance Use	–2.48*	.614***	–0.61	.743***	–1.07	.774***	–2.27*	.781***	–1.91	.787***	–1.62	.722***
9	Impulse Control	–5.16***	.563***	–0.41	.484***	–0.93	.525***	–0.16	.472***	–0.67	.499***	–3.17**	.515***
10	External Triggers	–2.87**	.584***	–1.85	.571***	–0.67	.569***	–0.12	.479***	–0.56	.426***	–0.74	.532***
11	Social Support	–1.23	.560***	–0.71	.635***	–0.15	.627***	–0.58	.620***	–1.72	.695***	–0.56	.617***
12	Material Resources	–0.24	.545***	–0.96	.603***	–1.89	.594***	–1.27	.578***	–0.58	.673***	–1.11	.591***
13	Attitudes	–4.43***	.602***	–0.28	.568***	–1.13	.705***	–0.16	.514***	–0.58	.561***	–3.01**	.592***
14	Medication Adherence	–2.61**	.330***	–1.23	.367***	–0.62	.428***	–1.91	.291**	–0.18	.491***	–0.35	.365***
15	Rule Adherence	–2.01*	.584***	–0.66	.604***	–0.37	.629***	–0.06	.653***	–0.24	.730***	–1.34	.630***
16	Conduct	–4.61***	.546***	–0.64	.627***	–1.26	.605***	–0.78	.589***	–0.85	.489***	–3.85***	.583***
17	Insight	–2.12*	.619***	–0.38	.578***	–1.40	.569***	–1.23	.610***	–0.52	.672***	–1.24	.604***
18	Plans	–1.66	.549***	–0.13	.658***	–0.41	.583***	–1.28	.570***	–0.33	.680***	–0.63	.602***
19	Coping	–3.48***	.368***	–1.15	.430***	–0.59	.509***	–0.42	.350***	–0.80	.507***	–1.59	.415***
20	Treatability	–0.11	.520***	–0.35	.455***	–0.80	.475***	–0.62	.560***	–0.68	.571***	–0.13	.511***
Strength		–4.49***	.813***	–0.29	.795***	–1.78	.807***	–0.24	.819***	–0.50	.807***	–2.96**	.813***
1	Social Skills	–2.26*	.466***	–0.78	.602***	–1.06	.655***	–0.17	.592***	0.00	.695***	–1.48	.587***

	START Rating	BL to FU1		FU1 to FU2		FU2 to FU3		FU3 to FU4		FU4 to FU5		Overall	
		Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s
2	Relationships	–3.20**	.575***	–1.44	.611***	–1.73	.599***	–0.16	.626***	–0.66	.566***	–1.46	.592***
3	Occupational	–0.66	.552***	–0.11	.585***	–0.12	.631***	–0.93	.672***	0.00	.707***	–0.73	.618***
4	Recreational	–0.22	.507***	–1.21	.568***	–2.02***	.538***	–0.28	.511***	–2.00*	.603***	–0.73	.545***
5	Self-Care	–0.23	.584***	–0.32	.708***	–0.69	.701***	–1.06	.674***	–0.01	.556***	–0.94	.645***
6	Mental State	–3.35***	.610***	–0.76	.383***	–1.33	.536***	–2.87**	.702***	–1.21	.633***	–2.40*	.564***
7	Emotional State	–3.02**	.453***	–0.81	.427***	–0.91	.514***	–0.85	.448***	–0.74	.530***	–1.52	.466***
8	Substance Use	–2.03*	.605***	–0.64	.670***	0.00	.668***	–1.23	.768***	–1.48	.682***	–1.32	.678***
9	Impulse Control	–2.09*	.498***	–0.41	.539***	–2.16*	.462***	–0.76	.534***	–0.48	.512***	–1.81	.522***
10	External Triggers	–1.60	.554***	–1.18	.617***	–1.02	.521***	–1.10	.549***	–0.61	.534***	–1.59	.568***
11	Social Support	–1.13	.620***	–0.67	.607***	–0.95	.675***	–2.60**	.771***	–0.23	.803***	–0.56	.681***
12	Material Resources	–0.93	.563***	–0.83	.501***	–0.20	.720***	–1.26	.624***	–0.34	.516***	–1.21	.587***
13	Attitudes	–3.35***	.642***	–0.89	.665***	–1.16	.566***	0.00	.486***	–0.97	.558***	–2.25*	.603***
14	Medication Adherence	–2.84**	.457***	–0.29	.451***	–1.27	.536***	–0.90	.649***	0.00	.614***	–0.61	.529***
15	Rule Adherence	–1.99*	.527***	–0.14	.579***	–0.49	.615***	–1.88	.704***	–2.12*	.696***	–1.29	.613***
16	Conduct	–3.13**	.512***	–0.85	.601***	–1.21	.604***	–0.30	.662***	–0.85	.521***	–2.41*	.578***
17	Insight	–3.05**	.681***	–1.24	.565***	–1.62	.644***	–0.52	.660***	–0.90	.650***	–1.04	.635***
18	Plans	–0.47	.570***	–1.31	.661***	–1.31	.604***	–1.00	.542***	–0.56	.710***	–0.01	.611***
19	Coping	–2.69**	.466***	–0.14	.434***	–1.18	.499***	0.00	.452***	–0.96	.574***	–2.35*	.484***
20	Treatability	–0.92	.521***	–0.28	.594***	–0.57	.570***	–2.03*	.609***	0.00	.654***	–0.01	.581***
Summary Risk Judgment		–3.00**	.652***	–0.14	.642***	–1.18	.674***	–0.73	.753***	–2.32*	.696***	–1.71	.671***

Note. Z = Z-coefficient from Wilcoxon signed-rank test (Z-coefficients are always negative); r_s = Spearman's rho stability coefficient; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3-20. Aggregate-Level Change on the START from Baseline

START Rating		BL to FU1		BL to FU2		BL to FU3		BL to FU4		BL to FU5	
		Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s
Vulnerability		–5.32***	.668***	–4.21***	.608***	–5.14***	.615***	–4.28***	.614***	–4.95***	.649***
1	Social Skills	–1.94	.456***	–2.03*	.412***	–2.90**	.505***	–2.00*	.476***	–2.59**	.467***
2	Relationships	–4.02***	.600***	–3.63***	.498***	–3.32***	.483***	–3.22**	.428***	–2.42*	.369***
3	Occupational	–1.54	.486***	–2.11*	.500***	–2.37*	.495***	–1.88	.370***	–2.29*	.456***
4	Recreational	–2.42*	.529***	–0.45	.475***	–0.74	.495***	–1.11	.311**	–0.26	.356***
5	Self-Care	–0.27	.524***	–1.07	.465***	–0.27	.386***	–0.59	.354***	–0.15	.339***
6	Mental State	–3.52***	.728***	–3.73***	.576***	–3.85***	.610***	–3.49***	.577***	–3.61***	.482***
7	Emotional State	–3.22**	.458***	–2.49*	.351***	–1.74	.415***	–1.60	.298**	–2.81**	.452***
8	Substance Use	–2.48*	.614***	–2.02*	.630***	–1.17	.604***	–2.21*	.584***	–1.15	.546***
9	Impulse Control	–5.16***	.563***	–4.66***	.607***	–4.66***	.491***	–4.82***	.553***	–4.72***	.489***
10	External Triggers	–2.87**	.584***	–1.16	.501***	–2.46*	.526***	–2.36*	.499***	–1.94	.497***
11	Social Support	–1.23	.560***	–1.39	.686***	–1.11	.633***	–1.21	.602***	–0.02	.445***
12	Material Resources	–0.24	.545***	–0.49	.475***	–1.76	.484***	–1.16	.513***	–1.27	.481***
13	Attitudes	–4.43***	.602***	–3.13**	.461***	–3.71***	.447***	–2.85**	.302**	–3.26**	.310**
14	Medication Adherence	–2.61**	.330***	–1.75	.292**	–2.40*	.169	–0.89	.116	–1.74	.335**
15	Rule Adherence	–2.01*	.584***	–2.04*	.608***	–1.94	.561***	–1.88	.508***	–2.05*	.534***
16	Conduct	–4.61***	.546***	–4.38***	.544***	–4.69***	.556***	–4.79***	.434***	–4.49***	.419***
17	Insight	–2.12*	.619***	–0.83	.461***	–2.00*	.508***	–0.82	.553***	–1.44	.428***
18	Plans	–1.66	.549***	–1.29	.441***	–1.54	.486***	–0.27	.389***	–1.38	.488***
19	Coping	–3.48***	.368***	–2.28*	.306***	–3.15**	.318***	–3.13**	.194*	–2.99**	.423***
20	Treatability	–0.11	.520***	–0.33	.377***	–0.71	.296**	–0.34	.343***	–0.80	.374***
Strength		–4.49***	.813***	–3.07**	.726***	–4.67***	.727***	–3.89***	.708***	–3.69***	.761***
1	Social Skills	–2.26*	.466***	–1.43	.493***	–1.70	.526***	–1.61	.433***	–2.00*	.620***

	START Rating	BL to FU1		BL to FU2		BL to FU3		BL to FU4		BL to FU5	
		Z	r_s	Z	r_s	Z	r_s	Z	r_s	Z	r_s
2	Relationships	–3.20**	.575***	–1.47	.489***	–3.33***	.565***	–2.38*	.510***	–2.14*	.571***
3	Occupational	–0.66	.552***	–0.91	.507***	–0.15	.541***	–0.98	.494***	–0.82	.537***
4	Recreational	–0.22	.507***	–1.06	.473***	–1.96	.477***	–1.52	.223*	–0.57	.469***
5	Self-Care	–0.23	.584***	–0.14	.564***	–0.56	.533***	–1.18	.514***	–1.04	.475***
6	Mental State	–3.35***	.610***	–1.14	.385***	–2.48*	.376***	–4.29***	.448***	–2.79**	.310**
7	Emotional State	–3.02**	.453***	–1.67	.279**	–2.39*	.381***	–1.66	.333***	–1.77	.317**
8	Substance Use	–2.03*	.605***	–1.69	.538***	–1.77	.618***	–2.44*	.535***	–1.17	.537***
9	Impulse Control	–2.09*	.498***	–2.83**	.543***	–4.13***	.424***	–3.31***	.359***	–3.16**	.436***
10	External Triggers	–1.60	.554***	–2.18*	.495***	–3.19**	.356***	–2.24*	.400***	–2.97**	.448***
11	Social Support	–1.13	.620***	–0.62	.555***	–1.94	.634***	–0.45	.599***	–0.16	.537***
12	Material Resources	–0.93	.563***	–1.34	.455***	–1.78	.477***	–3.00**	.467***	–1.77	.450***
13	Attitudes	–3.35***	.642***	–3.10**	.596***	–3.94***	.540***	–3.06**	.412***	–2.94**	.594***
14	Medication Adherence	–2.84**	.457***	–2.81**	.322***	–2.26*	.328**	–1.39	.353**	–1.31	.347**
15	Rule Adherence	–1.99*	.527***	–1.24	.442***	–1.61	.513***	–2.82**	.558***	–1.23	.483***
16	Conduct	–3.13**	.512***	–2.88**	.468***	–3.52***	.483***	–3.49***	.515***	–2.92**	.370***
17	Insight	–3.05**	.681***	–1.55	.543***	–2.34*	.563***	–1.73	.554***	–1.41	.585***
18	Plans	–0.47	.570***	–0.72	.538***	–0.62	.514***	–0.51	.443***	–0.26	.506***
19	Coping	–2.69**	.466***	–2.38*	.455***	–3.36***	.451***	–2.76**	.394***	–3.06**	.414***
20	Treatability	–0.92	.521***	–1.57	.556***	–1.62	.501***	–1.05	.489***	–0.30	.536***
Summary Risk Judgment		–3.00**	.652***	–2.31**	.571***	–3.91***	.668***	–3.88***	.538***	–3.04**	.592***

Note. Z = Z-coefficient from Wilcoxon signed-rank test (Z-coefficients are always negative); r_s = Spearman's rho stability coefficient; BL = baseline; FU1 = follow-up 1; FU2 = follow-up 2; FU3 = follow-up 3; FU4 = follow-up 4; FU5 = follow-up 5.
* $p < .05$, ** $p < .01$, *** $p < .001$.

With respect to the Wilcoxon tests, four items on the HCR-20 (i.e., Lack of Insight, Unresponsive to Treatment, Plans Lack Feasibility, and Lack of Personal Support) showed no mean rank change across any of the assessment intervals (see Table 3-17) or from the baseline to any follow-up (see Table 3-18). The remaining six items displayed mean rank change across some of the assessment intervals. More mean rank change was generally seen for these items earlier in the study period (i.e., baseline to follow-up one), whereas relatively few items experienced mean rank change from one follow-up to another (see Table 3-17). Three items (Active Symptoms of Major Mental Illness, Impulsivity, and Noncompliance with Treatment) experienced mean rank change in all analyses from baseline in Table 3-18. The summary risk judgments displayed change from baseline to each of the follow-up assessments (see Table 3-18), but not throughout the follow-up assessments (see Table 3-17). The scale score generally showed a similar pattern with mean rank change from baseline to each of the follow-ups, but not from one follow-up to the next.

All the stability coefficients were significant (all coefficients in Table 3-17 and all but three coefficients in Table 3-18 were significant at the $p < .001$ level). The stability coefficients for the risk factors varied from .41 to .83 across assessments and .23 to .71 from the baseline to each follow-up. The summary risk judgments yielded similar coefficients ranging from .57 to .79, while the scale scores ranged from .45 to .78. Based on Table 3-18, stability may decrease over time across all ratings on the HCR-20, as evidenced by the generally decreasing stability coefficients going from left to right (i.e., as the length of the assessment interval increases).

With respect to the Wilcoxon tests on the START, seven Vulnerability and five Strength items showed no mean rank change across any of the assessment intervals (see Table 3-19), while five Vulnerability and six Strength items showed no mean rank change from the baseline to any follow-up (see Table 3-20). Once again, more mean rank change was generally seen for the items earlier in the study period with 13 Vulnerability and 12 Strength items changing from baseline to the first follow-up, whereas relatively few items experienced mean rank change from one follow-up to the next (see Table 3-19). For instance, no items showed mean rank change from follow-up one to follow-up two and only two Strength items from follow-up two to follow-up three. No item experienced change in mean rank more than twice across the assessments. In contrast, six Vulnerability and four Strength items experienced mean rank change in all

analyses when compared from baseline in Table 3-20. The summary risk judgments displayed change from baseline to each of the follow-ups (see Table 3-20), as well as between the fourth and fifth follow-ups (see Table 3-19). The scale score showed a similar pattern with mean rank change from baseline to each of the follow-ups, but not from one follow-up to the next.

Nearly all the stability coefficients were significant (all but one coefficient in Table 3-19 and all but 16 coefficients in Table 3-20 were significant at the $p < .001$ level). The stability coefficients for the item ratings varied from .29 to .80 across each assessment and from .12 to .73 from the baseline to each follow-up. The summary risk judgments yielded similar coefficients ranging from .54 to .75, while the scale scores ranged from .61 to .82. Based on Table 3-20, a similar pattern was seen for the START as the HCR-20; stability may decrease over time across ratings on the START, as the stability coefficients generally decreased as the length of the assessment interval increased.

3.5.1. Characteristics of Internal Responsiveness

This section presents the results of the analyses examining the interrelationships amongst dynamic risk ratings over time. Specifically, to address research question 1a and describe some of the characteristics of internal responsiveness, repeated measures correlation was used to examine the relationship amongst ratings on the HCR-20 over time.[15] Repeated measures correlations were examined across all assessments for raw score ratings and change scores.

Table 3-21 displays the repeated measures correlations amongst all the dynamic ratings on the HCR-20 across assessments. Generally, the highest correlations were found between the Clinical (range = .50 to .60) and Risk Management (range = .56 to .63) scale scores and the respective items on that scale. Most items were correlated across assessments with most coefficients in the small range. The strongest item level relationship ($r = .40$) was seen between Unresponsive to Treatment (C5) and Noncompliance with Remediation Attempts (R4). Lack of Insight (C1) and Unresponsive to Treatment (R4) produced the next largest association ($r = .29$), followed by Lack of Personal Support (R3) and Stress (R5) ($r = .28$). All items and scale scores were

[15] Due to the exploratory nature of these analyses, only the HCR-20 was examined here.

correlated with the summary risk judgments (range = .16 to .40), with the largest coefficients found for both scale scores.

Table 3-22 displays the repeated measures correlations amongst change scores on all the dynamic ratings on the HCR-20 across assessments. Once again, the strongest correlations were found between change on the Clinical (range = .49 to .61) and Risk Management (range = .53 to .63) scale scores and change on items on the respective scale. Change on many items was correlated across assessments with most coefficients in the small range. The strongest item level relationship (r = .36) was seen between change on Unresponsive to Treatment (C5) and change on Noncompliance with Remediation Attempts (R4). Change on Lack of Insight (C1) and change on Unresponsive to Treatment (R4) produced the next largest association (r = .31), followed by change on Plans Lack Feasibility (R1) and change on Noncompliance with Remediation Attempts (R4) (r = .27). Change on some items was less associated with change on the other items and scales. For instance, Active Symptoms of Major Mental Illness (C3) produced some of the smallest coefficients (and most non-significant correlations). Change on all items and scale scores were correlated with change on the summary risk judgments (range = .14 to .36), with the largest coefficients found for both scale scores.

Table 3-21. Associations amongst HCR-20 Ratings across Assessments

HCR-20	C	C1	C2	C3	C4	C5	R	R1	R2	R3	R4	R5	SRJ
C Scale	1.00	.57***	.50***	.60***	.58***	.58***	.42***	.19***	.23***	.20***	.42***	.20***	.40***
C1		1.00	.13**	.15***	.13**	.25***	.27***	.20***	.13***	.09*	.29***	.09*	.20***
C2			1.00	.12**	.16***	.12**	.22***	.04	.18***	.08*	.23***	.11**	.22***
C3				1.00	.24***	.13***	.20***	.06	.10**	.11**	.14***	.15***	.21***
C4					1.00	.11**	.19***	.06	.11**	.14***	.15***	.10*	.27***
C5						1.00	.32***	.15***	.13***	.17***	.40***	.11**	.22***
R Scale							1.00	.56***	.63***	.58***	.58***	.59***	.40***
R1								1.00	.17***	.16***	.24***	.09*	.16***
R2									1.00	.22***	.23***	.24***	.26***
R3										1.00	.15***	.28***	.24***
R4											1.00	.10*	.27***
R5												1.00	.25***
SRJ													1.00

Note. Vales presented are repeated-measures correlation coefficients.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3-22. Associations amongst HCR-20 Change Scores across Assessments

HCR-20	C	C1	C2	C3	C4	C5	R	R1	R2	R3	R4	R5	SRJ
C Scale	1.00	.58***	.49***	.50***	.52***	.61***	.41***	.19***	.25***	.18***	.41***	.22***	.35***
C1		1.00	.13**	.06	.12**	.26***	.32***	.25***	.16***	.10*	.31***	.14**	.17***
C2			1.00	.09	.10*	.14**	.25***	.11*	.20***	.11*	.20***	.15**	.22***
C3				1.00	.11*	.10*	.09	-.04	.07	.05	.10*	.07	.14**
C4					1.00	.09	.17***	.04	.15**	.11*	.12**	.10*	.23***
C5						1.00	.30***	.14**	.11*	.14**	.36***	.13**	.21***
R Scale							1.00	.53***	.63***	.53***	.57***	.55***	.36***
R1								1.00	.11*	.14**	.27***	.06	.20***
R2									1.00	.22***	.23***	.22***	.25***
R3										1.00	.09	.18***	.24***
R4											1.00	.11*	.20***
R5												1.00	.23***
SRJ													1.00

Note. Vales presented are repeated-measures correlation coefficients.
* p < .05, ** p < .01, *** p < .001.

3.6. External Responsiveness

This section presents the results of the analyses of external responsiveness. Generalized estimating equations (GEE) was used to address the second research question concerning the relationship between change on the dynamic ratings and violence. First, the bivariate relationship between repeated assessments on each of the dynamic risk ratings on both tools and violence was examined. Next, the corresponding baseline rating was included as a covariate, to address the incremental validity of repeated assessments on these dynamic ratings over a single baseline assessment. Then, the HCR-20 Historical scale was included as a covariate, to address the incremental validity of repeated assessments over static risk factor scores. Finally, change scores were included as opposed to raw scores to examine the relationship between change on the dynamic ratings and future violence across assessments.

Table 3-23 presents the results of the GEE analyses for the HCR-20 for overall violence and Table 3-24 presents the results for the more narrow or serious definition of violence. With respect to the perpetration of any violence, four Clinical items (Lack of Insight, Negative Attitudes, Impulsivity, and Unresponsive to Treatment) and four Risk Management items (Plans Lack Feasibility, Exposure to Destabilizers, Lack of Personal Support, and Noncompliance with Remediation Attempts) were predictive of violence across assessments (*OR* = 1.90 to 3.10). When controlling for the respective baseline scores, each of these items, except Negative Attitudes, was still associated with future violence (*OR* = 1.69 to 2.31), and they were all still associated with violence after controlling for the Historical scale scores (*OR* = 1.48 to 2.48). One Risk Management item (Stress) was predictive of violence only when controlling for the baseline scores, while one Clinical item (Active Symptoms of Major Mental Illness) was not predictive of violence alone or when controlling for baseline scores.

The summary risk judgments were predictive of violence across repeated assessments (*OR* = 3.44), when controlling for baseline rating (*OR* = 3.58), and when controlling for static risk (*OR* = 2.53). Both the Clinical and Risk Management scale scores were also predictive of overall violence and when controlled for their baseline scores or the Historical scale. In contrast, when change from one assessment to the next was used to predict future violence, only a single item (Active Symptoms of Major Mental Illness) was predictive of violence.

Table 3-23. External Responsiveness of the HCR-20 – Overall Violence

HCR-20 Rating		Rating		Controlling Baseline		Controlling H Scale		Change Score	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Clinical (C) Scale		1.41***	[1.26, 1.59]	1.37***	[1.16, 1.61]	1.27***	[1.14, 1.42]	0.96	[0.84, 1.09]
C1	Lack of Insight	1.90***	[1.35, 2.68]	2.31***	[1.47, 3.63]	1.48*	[1.03, 2.13]	0.70	[0.45, 1.10]
C2	Negative Attitudes	2.72***	[1.87, 3.93]	1.48	[0.92, 2.39]	1.86**	[1.26, 2.74]	1.45	[0.93, 2.26]
C3	Active Symptoms of MMI	0.93	[0.53, 1.45]	1.66	[0.97, 2.86]	0.93	[0.61, 1.42]	0.67**	[0.50, 0.88]
C4	Impulsivity	2.98***	[2.16, 4.12]	2.05***	[1.43, 2.93]	2.17***	[1.59, 2.97]	0.97	[0.67, 1.41]
C5	Unresponsive to Tx	2.21***	[1.57, 3.06]	1.99***	[1.35, 2.93]	1.63**	[1.16, 2.29]	1.05	[0.78, 1.42]
Risk Management (R) Scale		1.42***	[1.27, 1.59]	1.31***	[1.12, 1.53]	1.29***	[1.15, 1.44]	0.98	[0.87, 1.10]
R1	Plans Lack Feasibility	3.10***	[2.16, 4.47]	2.25***	[1.36, 3.70]	2.48***	[1.76, 3.47]	1.03	[0.68, 1.56]
R2	Exposure to Destabilizers	2.68***	[1.80, 3.98]	2.07**	[1.30, 3.29]	1.84**	[1.24, 2.74]	1.02	[0.72, 1.46]
R3	Lack of Personal Support	2.12***	[1.48, 3.04]	1.99**	[1.26, 3.16]	1.55*	[1.06, 2.25]	0.83	[0.60, 1.14]
R4	Noncompliance w/ Tx	2.56***	[1.76, 3.72]	1.80**	[1.15, 2.81]	1.79**	[1.23, 2.61]	1.06	[0.69, 1.61]
R5	Stress	1.46	[0.95, 2.25]	1.69*	[1.02, 2.81]	1.32	[0.87, 2.01]	0.87	[0.62, 1.23]
Summary Risk Judgment		3.42***	[2.41, 4.84]	3.58***	[2.22, 5.78]	2.53***	[1.66, 3.86]	1.02	[0.59, 1.76]

Note. OR = odds ratio. CI = confidence interval.
* p < .05, ** p < .01, *** p < .001.

Table 3-24. External Responsiveness of the HCR-20 – Serious Violence

HCR-20 Rating		Rating		Controlling Baseline		Controlling H Scale		Change Score	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Clinical (C) Scale		1.41***	[1.23, 1.61]	1.32*	[1.03, 1.69]	1.21**	[1.07, 1.37]	0.90	[0.73, 1.11]
C1	Lack of Insight	1.99***	[1.36, 2.91]	2.52***	[1.45, 4.38]	1.46	[0.96, 2.23]	0.49*	[0.25, 0.94]
C2	Negative Attitudes	3.12***	[2.03, 4.80]	1.51	[0.85, 2.67]	1.96**	[1.26, 3.05]	1.40	[0.77, 2.54]
C3	Active Symptoms of MMI	0.93	[0.59, 1.45]	1.66	[0.97, 2.86]	0.93	[0.61, 1.42]	0.66**	[0.50, 0.88]
C4	Impulsivity	3.41***	[2.22, 5.32]	1.96**	[1.19, 3.23]	2.28***	[1.51, 3.43]	0.92	[0.59, 1.41]
C5	Unresponsive to Tx	1.97***	[1.36, 2.86]	1.55	[0.91, 2.26]	1.31	[0.87, 1.99]	1.13	[0.74, 1.73]
Risk Management (R) Scale		1.44***	[1.26, 1.64]	1.24*	[1.03, 1.51]	1.25***	[1.09, 1.43]	0.98	[0.87, 1.10]
R1	Plans Lack Feasibility	3.18***	[2.02, 5.00]	2.09*	[1.12, 3.91]	2.36***	[1.57, 3.54]	0.89	[0.56, 1.43]
R2	Exposure to Destabilizers	3.39***	[2.05, 5.62]	2.47***	[1.50, 4.07]	2.11***	[1.40, 3.17]	0.89	[0.68, 1.16]
R3	Lack of Personal Support	2.26***	[142, 3.61]	1.57	[0.87, 2.82]	1.49	[0.85, 2.59]	1.21	[0.78, 1.89]
R4	Noncompliance w/ Tx	2.74***	[1.79, 4.19]	1.62	[0.88, 2.97]	1.74*	[1.08, 2.79]	1.19	[0.70, 2.03]
R5	Stress	1.26	[0.73, 2.15]	1.49	[0.79, 2.82]	1.08	[0.67, 1.76]	0.83	[0.58, 1.19]
Summary Risk Judgment		4.24***	[2.85, 6.30]	4.33***	[2.30, 8.16]	3.00***	[1.70, 5.31]	0.89	[0.42, 1.91]

Note. OR = odds ratio. CI = confidence interval.
* p < .05, ** p < .01, *** p < .001.

With respect to the perpetration of serious violence, the same four Clinical items (Lack of Insight, Negative Attitudes, Impulsivity, and Unresponsive to Treatment) and four Risk Management items (Plans Lack Feasibility, Exposure to Destabilizers, Lack of Personal Support, and Noncompliance with Remediation Attempts) were predictive of violence (*OR* = 1.97 to 3.41). However, when the respective baseline scores were controlled for in the models, only two items from each scale (Lack of Insight, Impulsivity, Plans Lack Feasibility, and Exposure to Destabilizers) were still associated with violence (*OR* = 1.96 to 2.52). When controlling for Historical scale, two Clinical items and three Risks Management items were associated with violence (*OR* = 1.74 to 2.36). One Clinical item (Active Symptoms of Major Mental Illness) and one Risk Management item (Stress) were not predictive of violence alone or when controlling for baseline scores.

The summary risk judgments were predictive of serious violence independently (*OR* = 4.24), when controlling for the baseline risk (*OR* = 4.33), and when controlling for the Historical scale (*OR* = 3.00). Both the Clinical and Risk Management scale scores were also predictive of serious violence alone and when controlling for their baseline scores or the Historical scale. In contrast, when change scores were analyzed, only two items (Lack of Insight and Active Symptoms of Major Mental Illness) were predictive of serious violence (*OR* = .49 and .66, respectively).

Overall, the strongest relationships were seen for the summary risk judgments for general and serious violence. In each instance, an increase in summary risk rating corresponded with an approximate three-fold or four-fold increase in the odds of violence, respectively. Several items also demonstrated rather stable relationships with violence controlling for baseline scores or static risk scores. An increase in the presence of these items corresponded with an approximate two-fold increase in the odds of violence. Although the summary risk judgments tended to yield larger odds ratios for serious violence, the scale scores showed a relatively stable association with overall and serious violence (i.e., the odds ratios varied minimally across analyses).

Table 3-25 presents the results of the GEE analyses for the START for overall violence and Table 3-26 presents the results for the more narrow or serious definition of violence. With respect to the perpetration of any violence, 16 Vulnerability items (*OR* = 1.61 to 3.57) and 18 Strength items (*OR* = 0.33 to 0.63) were predictive of violence. When controlling for the respective baseline scores, 12 Vulnerability (*OR* = 1.49 to 2.42)

and ten Strength items (*OR* = 0.41 to 0.62) were still associated with future violence. As well, 11 Vulnerability (*OR* = 1.66 to 2.56) and 10 Strength items (*OR* = 0.44 to 0.71) were associated with violence after controlling for the Historical scale scores. Notably, four Vulnerability items (Recreational, Mental State, Emotional State, and Insight) and one Strength item (Emotional State) were not predictive of violence alone or when controlling for any baseline scores.

The summary risk judgments were predictive of violence independently (*OR* = 3.71), when controlling for the baseline risk (*OR* = 2.56), and when controlling for the Historical scale (*OR* = 2.74). Both the Vulnerability and Strength scale scores were also predictive of overall violence alone and when controlling for their baseline scores or the Historical scale. Like the HCR-20, when change from one assessment to the next was used to predict future violence, only a single item on the Vulnerability scale (Self-Care) was predictive of violence.

With respect to the perpetration of serious violence, 12 Vulnerability items (*OR* = 1.81 to 3.88) and 16 Strength items (*OR* = 0.23 to 0.53) were predictive of violence. When controlling for the baseline ratings, 12 Vulnerability (*OR* = 1.25 to 2.38) and seven Strength items (*OR* = 0.33 to 0.62) were predictive of violence. As well, 10 Vulnerability (*OR* = 1.40 to 2.61) and nine Strength items (*OR* = 0.31 to 0.59) were associated with violence after controlling for the Historical scale. Overall, four Vulnerability items (Social Skills, Mental State, Emotional State, and Coping) and four Strength item (Self-Care, Mental State, Emotional State, and Material Resources) were not predictive of violence alone or when controlling for baseline scores.

The summary risk judgments were predictive of serious violence independently (*OR* = 4.16), when controlling for the baseline risk (*OR* = 2.65), and when controlling for the Historical scale (*OR* = 2.88). The Vulnerability scale scores were also predictive of serious violence alone and when controlling for their baseline scores or the Historical scale, while the Strength scale scores were predictive of serious violence alone and when controlling for the Historical scale. Unlike the number of associations reported above, when change scores were examined, a single item on the Vulnerability scale (Material Resources) was predictive of serious violence.

Table 3-25. External Responsiveness of the START – Overall Violence

START Rating		Rating		Controlling Baseline		Controlling H Scale		Change Score	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Vulnerability		1.12***	[1.09, 1.16]	1.09**	[1.03, 1.15]	1.10***	[1.06, 1.14]	0.97	[0.94, 1.00]
1	Social Skills	1.77**	[1.19, 2.64]	1.21	[0.78, 1.85]	1.47	[0.93, 2.30]	1.14	[0.84, 1.56]
2	Relationships	3.44***	[2.35, 5.04]	2.05***	[1.34, 3.13]	2.56***	[1.72, 3.80]	0.86	[0.58, 1.28]
3	Occupational	1.76*	[1.02, 3.03]	1.59	[0.82, 3.10]	1.36*	[1.04, 1.80]	.093	[0.83, 1.04]
4	Recreational	1.46	[0.91, 2.34]	1.37	[0.91, 2.08]	1.19	[0.86, 1.65]	0.84	[0.71, 1.01]
5	Self-Care	1.61**	[1.13, 2.29]	1.92***	[1.30, 2.83]	1.37	[0.97, 1.93]	0.56**	[0.36, 0.87]
6	Mental State	1.12	[0.79, 1.58]	1.51	[0.94, 2.44]	1.20	[0.85, 1.70]	0.77	[0.52, 1.15]
7	Emotional State	1.27	[0.90, 1.77]	1.28	[0.89, 1.82]	1.20	[0.84, 1.72]	0.99	[0.79, 1.25]
8	Substance Use	2.05***	[1.48, 2.83]	1.49*	[1.02, 2.17]	1.47	[0.99, 2.18]	0.86	[0.58, 1.25]
9	Impulse Control	2.93***	[2.10, 4.10]	1.77**	[1.20, 2.62]	2.09***	[1.49, 2.92]	0.87	[0.64, 1.19]
10	External Triggers	2.74***	[1.87, 4.03]	1.73*	[1.11, 2.70]	1.94***	[1.38, 2.71]	0.90	[0.64, 1.26]
11	Social Support	2.46***	[1.62, 3.73]	2.33***	[1.48. 3.65]	1.84**	[1.18, 2.88]	0.74	[0.55, 1.01]
12	Material Resources	2.38***	[1.55, 3.65]	1.83*	[1.06, 3.14]	1.66*	[1.11, 2.48]	0.80	[0.55, 1.18]
13	Attitudes	3.57***	[2.40, 5.32]	2.21**	[1.38, 3.56]	2.62***	[1.79, 3.85]	1.02	[0.69, 1.52]
14	Medication Adherence	1.75**	[1.24, 2.47]	1.81**	[1.23, 2.66]	1.31	[0.93, 1.84]	.077	[0.56, 1.08]
15	Rule Adherence	2.35***	[1.70, 3.25]	1.53	[0.98, 2.38]	1.64**	[1.13, 2.39]	1.01	[0.69, 1.47]
16	Conduct	2.97***	[2.11, 4.15]	1.79**	[1.16, 2.76]	2.19***	[1.51, 3.19]	1.29	[0.80, 2.10]
17	Insight	1.57	[0.96, 2.57]	1.48	[0.91, 2.42]	1.24	[0.84, 1.85]	1.05	[0.93, 1.19]
18	Plans	2.66***	[1.86, 3.80]	2.35***	[1.62, 3.43]	2.16***	[1.51, 3.09]	0.98	[0.72, 1.34]
19	Coping	1.64*	[1.10, 2.44]	1.40	[0.93, 2.12]	1.48	[0.99, 2.22]	1.07	[0.78, 1.46]
20	Treatability	2.39***	[1.65, 3.46]	2.42***	[1.53, 3.85]	1.77**	[1.22, 2.56]	0.73	[0.49, 1.09]
Strength		0.92***	[0.89, 0.95]	0.94*	[0.89, 0.99]	0.94**	[0.91, 0.98]	1.01	[0.98, 1.05]

	START Rating	Rating		Controlling Baseline		Controlling H Scale		Change Score	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
1	Social Skills	0.43***	[0.29, 0.65]	0.60	[0.35, 1.01]	0.47***	[0.31, 0.72]	1.10	[0.70, 1.74]
2	Relationships	0.46***	[0.29, 0.73]	0.88	[0.53, 1.47]	0.61	[0.37, 1.00]	0.72	[0.48, 1.08]
3	Occupational	0.40***	[0.28, 0.59]	0.41***	[0.26, 0.63]	0.51**	[0.33, 0.78]	1.10	[0.91, 1.33]
4	Recreational	0.54***	[0.38, 0.76]	0.51**	[0.31, 0.83]	0.67*	[0.46, 0.98]	1.07	[0.86, 1.34]
5	Self-Care	0.63**	[0.45, 0.89]	0.56*	[0.34, 0.91]	0.71*	[0.51, 0.99]	1.26	[0.79, 2.01]
6	Mental State	0.81	[0.55, 1.20]	0.59*	[0.36, 0.98]	0.77	[0.53, 1.12]	1.34	[0.93, 1.94]
7	Emotional State	0.76	[0.52, 1.12]	0.66	[0.43, 1.00]	0.72	[0.46, 1.12]	1.25	[0.95, 1.65]
8	Substance Use	0.55***	[0.40, 0.76]	0.71	[0.50, 1.00]	0.75	[0.52, 1.08]	0.96	[0.68, 1.34]
9	Impulse Control	0.33***	[0.12, 0.52]	0.61	[0.36, 1.03]	0.44***	[0.26, 0.72]	0.90	[0.66, 1.21]
10	External Triggers	0.42***	[0.29, 0.62]	0.61**	[0.43, 0.86]	0.56**	[0.38, 0.84]	1.07	[0.82, 1.39]
11	Social Support	0.44***	[0.30, 0.66]	0.42***	[0.26, 0.69]	0.56**	[0.37, 0.87]	1.28	[0.92, 1.77]
12	Material Resources	0.65*	[0.42, 0.99]	0.81	[0.48, 1.37]	0.84	[0.53, 1.33]	0.81	[0.54, 1.21]
13	Attitudes	0.38***	[0.26, 0.58]	0.51**	[0.30, 0.84]	0.49**	[0.32, 0.77]	1.16	[0.80, 1.69]
14	Medication Adherence	0.60**	[0.41, 0.87]	0.57*	[0.34, 0.94]	0.75	[0.49, 1.16]	1.03	[0.68, 1.55]
15	Rule Adherence	0.47***	[0.33, 0.69]	0.62*	[0.42, 0.91]	0.66	[0.41, 1.05]	1.04	[0.74, 1.44]
16	Conduct	0.46***	[0.31, 0.70]	0.77	[0.48, 1.24]	0.67	[0.41, 1.08]	0.74	[0.48, 1.14]
17	Insight	0.62*	[0.43, 0.91]	0.85	[0.53, 1.35]	0.76	[0.47, 1.23]	0.84	[0.56, 1.27]
18	Plans	0.43***	[0.29, 0.64]	0.59	[0.34, 1.01]	0.51***	[0.34, 0.77]	1.10	[0.80, 1.51]
19	Coping	0.57*	[0.36, 0.92]	0.72	[0.42, 1.25]	0.69	[0.41, 1.17]	0.81	[0.57, 1.15]
20	Treatability	0.44***	[0.29, 0.66]	0.54**	[0.35, 0.84]	0.57*	[0.36, 0.89]	1.05	[0.75, 1.48]
Summary Risk Judgment		3.71***	[2.63, 5.24]	2.56***	[1.66, 3.95]	2.74***	[1.79, 4.18]	1.02	[0.67, 1.56]

Note. OR = odds ratio. CI = confidence interval.
* p < .05, ** p < .01, *** p < .001.

Table 3-26. External Responsiveness of the START – Serious Violence

START Rating		Rating		Controlling Baseline		Controlling H Scale		Change Score	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Vulnerability		1.12***	[1.08, 1.16]	1.06*	[1.01, 1.12]	1.08***	[1.03, 1.13]	0.98	[0.95, 1.01]
1	Social Skills	1.52	[0.95, 2.42]	0.97	[0.55, 1.70]	1.15	[0.67, 1.98]	1.31	[0.88, 1.96]
2	Relationships	3.88***	[2.43, 6.17]	1.88*	[1.10, 3.21]	2.61***	[1.58, 4.32]	1.06	[0.67, 1.68]
3	Occupational	1.75	[0.87, 3.53]	1.50	[0.85, 2.65]	1.40**	[1.09, 1.80]	0.93	[0.82, 1.04]
4	Recreational	1.28	[0.99, 1.63]	1.25*	[1.00, 1.55]	1.02	[0.78, 1.35]	0.87	[0.75, 1.02]
5	Self-Care	1.44	[0.89, 2.33]	1.81*	[1.15, 2.85]	1.13	[0.74, 1.72]	0.75	[0.45, 1.28]
6	Mental State	0.93	[0.62, 1.38]	1.22	[0.72, 2.05]	0.98	[0.68, 1.42]	0.83	[0.50, 1.35]
7	Emotional State	1.16	[0.84, 1.59]	1.16	[0.85, 1.57]	1.06	[0.80, 1.41]	1.05	[0.79, 1.41]
8	Substance Use	2.66***	[1.79, 3.96]	1.55	[0.99, 2.40]	1.82*	[1.07, 3.11]	0.92	[0.63, 1.35]
9	Impulse Control	3.60***	[2.31, 5.61]	2.02**	[1.24, 3.29]	2.35***	[1.49, 3.72]	0.86	[0.65, 1.13]
10	External Triggers	3.78***	[2.26, 6.32]	2.07*	[1.19, 3.61]	2.49***	[1.71, 3.62]	0.76	[0.51, 1.13]
11	Social Support	2.66***	[1.54, 4.59]	1.76*	[1.05, 2.96]	1.80	[0.95, 3.43]	0.95	[0.69, 1.31]
12	Material Resources	2.94***	[1.77, 4.88]	2.38**	[1.25, 4.50]	1.89**	[1.18, 3.01]	0.60*	[0.39, 0.91]
13	Attitudes	3.15***	[1.96, 5.07]	1.67	[0.97, 2.88]	1.97**	[1.28, 3.03]	1.37	[0.87, 2.15]
14	Medication Adherence	1.81**	[1.18, 2.78]	2.01*	[1.17, 3.44]	1.20	[0.79, 1.81]	0.73	[0.50, 1.07]
15	Rule Adherence	2.74***	[1.93, 3.89]	1.65	[0.98, 2.75]	1.80*	[1.11, 2.92]	0.99	[0.69, 1.43]
16	Conduct	3.07***	[2.08, 4.53]	1.63*	[1.04, 2.56]	2.08**	[1.29, 3.33]	1.51	[0.98, 2.32]
17	Insight	1.57	[0.90, 2.73]	1.40*	[1.04, 1.88]	1.23	[0.86, 1.75]	1.05	[0.90, 1.22]
18	Plans	2.43***	[1.57, 3.76]	1.85*	[1.13, 3.06]	1.83**	[1.16, 2.87]	0.95	[0.68, 1.34]
19	Coping	1.49	[0.90, 2.45]	1.38	[0.82, 2.33]	1.26	[0.78, 2.03]	0.91	[0.63, 1.30]
20	Treatability	2.12***	[1.45, 3.10]	1.90*	[1.08, 3.34]	1.40	[0.88, 2.23]	0.79	[0.49, 1.28]
Strength		0.91***	[0.87, 0.95]	0.96	[0.90, 1.02]	0.93**	[0.88, 0.98]	1.01	[0.96, 1.06]

	START Rating	Rating		Controlling Baseline		Controlling H Scale		Change Score	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
1	Social Skills	0.49**	[0.29, 0.83]	0.74	[0.37, 1.50]	0.58	[0.33, 1.02]	0.97	[0.59, 1.59]
2	Relationships	0.37***	[0.21, 0.66]	0.89	[0.48, 1.66]	0.54	[0.28, 1.06]	0.74	[0.47, 1.18]
3	Occupational	0.33***	[0.21, 0.53]	0.33***	[0.21, 0.54]	0.44**	[0.25, 0.77]	1.10	[0.91, 1.34]
4	Recreational	0.45***	[0.29, 0.69]	0.42***	[0.25, 0.70]	0.59*	[0.36, 0.97]	1.13	[0.81, 1.56]
5	Self-Care	0.70	[0.46, 1.06]	0.63	[0.37, 1.07]	0.83	[0.55, 1.24]	1.07	[0.64, 1.78]
6	Mental State	0.91	[0.54, 1.52]	0.72	[0.39, 1.33]	0.88	[0.55, 1.40]	1.21	[0.72, 2.00]
7	Emotional State	0.81	[0.53, 1.26]	0.87	[0.56, 1.36]	0.78	[0.47, 1.30]	1.23	[0.95, 1.58]
8	Substance Use	0.40***	[0.28, 0.58]	0.62*	[0.41, 0.93]	0.56**	[0.36, 0.87]	0.96	[0.65, 1.42]
9	Impulse Control	0.23***	[0.14, 0.40]	0.50*	[0.28, 0.93]	0.31***	[0.16, 0.60]	0.91	[0.67, 1.23]
10	External Triggers	0.38***	[0.24, 0.60]	0.77	[0.52, 1.13]	0.55*	[0.32, 0.93]	0.87	[0.70, 1.09]
11	Social Support	0.37***	[0.22, 0.61]	0.41**	[0.23, 0.72]	0.50*	[0.27, 0.92]	0.94	[0.70, 1.40]
12	Material Resources	0.64	[0.39, 1.05]	0.82	[0.42, 1.58]	0.86	[0.48, 1.54]	0.87	[0.53, 1.43]
13	Attitudes	0.31***	[0.19, 0.50]	0.43**	[0.24, 0.78]	0.41**	[0.23, 0.74]	1.26	[0.81, 1.98]
14	Medication Adherence	0.53**	[0.34, 0.85]	0.50	[0.23, 1.08]	0.74	[0.41, 1.32]	1.07	[0.64, 1.80]
15	Rule Adherence	0.33***	[0.22, 0.50]	0.52**	[0.32, 0.83]	0.45*	[0.25, 0.83]	1.01	[0.67, 1.55]
16	Conduct	0.37***	[0.23, 0.59]	0.69	[0.41, 1.14]	0.56	[0.29, 1.06]	0.70	[0.42, 1.15]
17	Insight	0.49***	[0.32, 0.76]	0.76	[0.41, 1.43]	0.59	[0.32, 1.09]	1.14	[0.64, 2.03]
18	Plans	0.47**	[0.29, 0.75]	0.69	[0.39, 1.23]	0.60	[0.34, 1.03]	1.18	[0.86, 1.62]
19	Coping	0.50*	[0.28, 0.91]	0.87	[0.45, 1.69]	0.64	[0.31, 1.28]	0.78	[0.48, 1.25]
20	Treatability	0.38***	[0.23, 0.61]	0.57	[0.32, 1.01]	0.52*	[0.28, 0.95]	0.94	[0.59, 1.51]
Summary Risk Judgment		4.16***	[2.83, 6.11]	2.65***	[1.56, 4.52]	2.88***	[1.58, 5.25]	1.08	[0.59, 1.98]

Note. OR = odds ratio. CI = confidence interval.
* p < .05, ** p < .01, *** p < .001.

Overall, the strongest relationships were generally seen for the summary risk judgments for general and serious violence. In each instance, an increase in summary risk rating corresponded with an approximate two-fold to four-fold increase in the odds of violence, depending on the analysis. Again, various items demonstrated a relationship with violence after controlling for baseline scores or static risk scores. Like the HCR-20, the summary risk judgments tended to yield larger odds ratios for serious violence, whereas the scale scores showed a relatively stable association with both definitions of violence (i.e., the odds ratios varied minimally across analyses).

3.7. Comparative Responsiveness

This section presents the results of the comparative responsiveness analyses. The proportions of participants who changed on the dynamic risk ratings (raw score and reliable change) were compared using the test developed by McNemar (1947). Additionally, the stability coefficients were compared using the test developed by Raghunathan and colleagues (1996). These tests focused on the total number of assessment pairs (N = 623).

The scale scores and summary risk judgments were first compared within and across instruments. On the HCR-20, there was no difference in the proportions of participants who changed on the Clinical and Risk Management scales (74.0% and 73.3%, p = .844), nor in the proportions showing reliable change (1.4% and 1.3%, p = 1.00). No difference was seen in rank order stability coefficients for the Clinical (r_s = .713) and Risk Management (r_s = .696) scales, Z = 0.69, p = .485. Compared to the summary risk judgments (21.5%), more change was seen on both the Clinical (74.0%) and Risk Management (73.3%) scales (both p < .001); however, no differences were seen in the proportions of reliable change for these ratings. As well, the stability coefficients did not vary between the summary risk judgments (r_s = .692) and the Clinical scale, Z = 0.85, p = 0.395, or the Risk Management scale, Z = 0.16, p = .875.

On the START, there was no difference in the proportions of participants who changed on the Vulnerability and Strength scales (90.0 and 91.5%, p = .426), yet more reliable change was seen on the Vulnerability scale compared to the Strength scale (4.4% and 0.6%, p < .001). As well, the stability coefficients differed between the Vulnerability and Strength scales (Z = 4.36, p < .001). Compared to the summary risk

judgments (19.7%), more change was seen on both the Vulnerability (90.0%) and Strength (91.5%) scales (both $p < .001$). More reliable change was also seen on the Vulnerability scale (4.4%) compared to the summary risk judgments (1.6%), but no difference was seen between the summary risk judgments and the Strength scale ($p = .146$). The stability coefficients also differed between the summary risk judgments and the Vulnerability scale ($Z = –2.65$, $p = .008$) and the Strength scale ($Z = –6.15$, $p < .001$).

Across tools, no difference was found between the proportions of participants experiencing change on the HCR-20 summary risk judgments compared to the START summary risk judgments (21.5% and 19.7%, $p = .169$). The same pattern was also seen for reliable change on these ratings (1.1% and 1.6%, $p = .375$), as well as the stability coefficients of these ratings ($Z = 1.27$, $p = .203$). Comparatively, the START scales experienced greater raw score change compared to the HCR-20 scales. That is, the Vulnerability scale changed more often than the Clinical scale ($p < .001$) or the Risk Management scale ($p < .001$). The Strength scale also changed more often than the Clinical scale ($p < .001$) or the Risk Management scale ($p <.001$). Looking at reliable change, this pattern was also seen for the Vulnerability scale but not the Strength scale. The Vulnerability scale experienced more reliable change than the Clinical ($p = .002$) or the Risk Management scales ($p < .001$), but the Strength scale did not experience more reliable change than the Clinical ($p = .227$) or the Risk Management scale ($p = .344$). Additionally, the stability coefficients differed between the Strength scale and both the Clinical ($Z = –4.89$, $p < .001$) and Risk Management ($Z = –5.61$, $p < .001$) scales, whereas stability coefficients for the Vulnerability scale differed from the Risk Management scale ($Z = –2.01$, $p = .044$) but not the Clinical scale ($Z = –1.19$, $p = .235$).

Next, specific post-hoc analyses were performed across select ratings on the HCR-20. The proportions of change and reliable change, as well as the stability coefficients, for the summary risk judgments and four items were examined. The items were selected based on observed lower (Negative Attitudes and Lack of Personal Support) and higher (Unresponsive to Treatment and Stress) rates of change (see Tables 3-9, 3-11, and 3-13). These analyses continued to address the comparative responsiveness of these ratings. Table 3-27 presents the results of these post-hoc comparisons on the HCR-20. In this table, the first two columns present the percent of change and reliable change, while the remaining columns present the results of the McNemar tests for the selected ratings.

Table 3-27. Comparative Responsiveness of Ratings on the HCR-20

HCR-20 Rating	Change		SRJ		C2		C5		R3		R5	
	C	RC	C	RC	C	RC	C	RC	C	RC	C	RC
Clinical (C) Scale	74.0	1.4	**<.001**[a]	.791[b]	**<.001**[a]	.791[b]	**<.001**[a]	**.001**[a]	**<.001**[a]	.549[b]	**<.001**[a]	.230[b]
C1 Lack of Insight	35.2	1.0	**<.001**[a]	1.00[b]	**<.001**[a]	.332[b]	.951[a]	**<.001**[a]	**.009**[a]	1.00[b]	.055[a]	**.041**[b]
C2 Negative Attitudes	25.4	1.8	.099[a]	.454[b]	-	-	**<.001**[a]	**.009**[a]	.299[a]	.302[b]	**<.001**[a]	.424[b]
C3 Active Symptoms of MMI	31.4	0.0	**<.001**[a]	**.016**[b]	**.027**[a]	**<.001**[b]	.128[a]	**<.001**[a]	.238[a]	**.031**[b]	**.001**[a]	**<.001**[b]
C4 Impulsivity	33.4	2.4	**<.001**[a]	.134[b]	**.002**[a]	.556[a]	.474[b]	.061[a]	**.042**[a]	.078[b]	**.012**[a]	1.00[a]
C5 Unresponsive to Tx	35.5	4.5	**<.001**[a]	**<.001**[a]	**<.001**[a]	**.009**[a]	-	-	**.006**[a]	**<.001**[a]	.082[a]	.090[a]
Risk Management (R) Scale	73.3	1.3	**<.001**[a]	1.00[b]	**<.001**[a]	.629[b]	**<.001**[a]	**<.001**[a]	**<.001**[a]	.754[b]	**<.001**[a]	.115[b]
R1 Plans Lack Feasibility	34.3	2.7	**<.001**[a]	0.64[b]	**<.001**[a]	.345[a]	.707[a]	.136[a]	**.018**[a]	**.035**[b]	**.030**[a]	1.00[a]
R2 Exposure to Destabilizers	37.7	3.0	**<.001**[a]	0.12[b]	**<.001**[a]	.201[a]	.427[a]	.212[a]	**<.001**[a]	**.011**[b]	.352[a]	.719[a]
R3 Lack of Personal Support	28.2	1.0	**.008**[a]	1.00[b]	.299[a]	.302[b]	**.006**[a]	**<.001**[a]	-	-	**<.001**[a]	**.041**[b]
R4 Noncompliance w/ Tx	34.7	1.3	**<.001**[a]	1.00[b]	**<.001**[a]	.648[b]	.789[a]	**<.001**[a]	**.014**[b]	.791[b]	**.038**[a]	.115[b]
R5 Stress	40.3	2.6	**<.001**[a]	.078[b]	**<.001**[a]	.424[b]	.082[a]	.090[a]	**<.001**[a]	**.041**	-	-
Summary Risk Judgment	21.5	1.1	-	-	.099[a]	.454[b]	**<.001**[a]	**<.001**[a]	**.008**[a]	1.00	**<.001**[a]	.078[b]

Note. Values presented in the first two columns are valid percents, while the remaining columns are [b]exact or [a]asymptotic p-values from McNemar's tests. Significant values are bolded. C = raw score change; RC = reliable change.

The summary risk judgments changed in fewer assessments than nine of ten dynamic risk factors and both scale scores. For reliable change, the summary risk judgments changed less than Unresponsive to Treatment and more than Active Symptoms of Major Mental Illness. Negative Attitudes changed in fewer assessments than all other items except Lack of Personal Support, yet it only experienced less reliable change than Unresponsive to Treatment and more than Active Symptoms of Major Mental Illness. Unresponsive to Treatment changed more often than two other items but experienced more reliable change than five other items and both scale scores. Lack of Personal Support experienced less frequent change than three Clinical items and all other Risk Management items, and less reliable change than Unresponsive to Treatment, Plans Lack Feasibility, Exposure to Destabilisers, and Stress. Stress changed more often than three Clinical and three Risk Management items and experienced more reliable change than three items (Lack of Insight, Active Symptoms of Major Mental Illness, and Lack of Personal Support).

As well, select comparisons were made across the stability coefficients of the select four items. The stability of Negative Attitudes (r_s = .641) was lower than the stability of Lack of Personal Support (r_s = .733), Z = –3.18, p = .002, as well as higher than the stability of Unresponsive to Treatment (r_s = .507), Z = 3.74, p < .001, and Stress (r_s = .470), Z = 4.42, p < .001. The stability of Unresponsive to Treatment was also lower than the stability of Lack of Personal Support, Z = –6.92, p < .001, whereas no difference was found between the stability of Unresponsive to Treatment and Stress, Z = 0.86, p = .388. As well, the stability of Lack of Personal Support was higher than the stability of Stress, Z = 7.72, p < .001.

The rank order stability of the summary risk judgments was also compared to the stability of all items. The stability of the summary risk judgments (r_s = .692) was higher than the stability of three Clinical items and four Risk Management items. On the Clinical scale, Lack of Insight (r_s = .631), Z = 2.03, p = .042, Impulsivity (r_s = .611), Z = 2.77, p = .006, and Unresponsive to Treatment (r_s = .507), Z = 5.53, p < .001, had lower stability than the summary risk judgments. On the Risk Managements scale, Plans Lack Feasibility (r_s = .628), Z = 2.13, p = .034, Exposure to Destabilizers (r_s = .592), Z = 3.40, p < .001, Noncompliance with Remediation Attempts (r_s = .595), Z = 3.25, p = .001, and Stress (r_s = .470), Z = 6.18, p < .001, had lower stability. The other three items showed no difference in stability compared to the summary risk judgments: Negative Attitudes (r_s

= .641), Z = 1.84, p = .065, Active Symptoms of Major Mental Illness (r_s = .700), Z = –0.27, p = .784, and Lack of Personal Support (r_s = .733), Z = –1.57, p = .118.

In addition, potential differences in the number of times participants changed on the select items and summary risk judgments were examined using bivariate McNemar Tests. That is, these tests examined the proportions of people who changed zero, one, two, three, four, or five times across the assessment intervals on each of the ratings as presented in Table 3-11. These results are generally consistent with the analyses above.

Comparing the summary risk judgments to each of the select items, participants were more likely to change once (p = .004) or twice (p = .018) on Unresponsive to Treatment and less likely have changed zero times (p < .001). Larger proportions of the sample also changed twice (p < .001) or four times (p = .049) on the item Stress and a smaller proportion changed no times (p < .001) compared to the summary risk judgments. In contrast, although a difference was seen between the proportions of participants who did not change on the summary risk judgments and Lack of Personal Support (p = .040), no other differences were seen in the proportions who changed once, twice, or more times. As well, no differences were found in the number of times changed when comparing the summary risk judgments and Negative Attitudes.

Comparing across the select items, several differences were identified. Unresponsive to Treatment was more likely to have changed once (p = .016) or twice (p = .045) and less likely to have experienced no change (p < .001) compared to Negative Attitudes. Stress was also more likely to have changed twice (p < .001) and less likely to have changed zero times (p < .001) compared to Negative Attitudes. On the other hand, the number of times people changed did not differ between Negative Attitudes and Lack of Personal Support. Participants were more likely to have changed once (p = .026) on Unresponsive to Treatment and less likely to have changed no times (p = .003) compared to Lack of Personal Support. Participants were also more likely to have changed twice (p = .015) on Stress and less likely to have changed no times (p < .001) compared to Lack of Personal Support. Finally, the number of times people changed across assessments did not differ between Unresponsive to Treatment and Stress.

Chapter 4. Discussion

It is common practice for a comprehensive violence risk assessment to focus on the assessment of risk state based on the consideration of a combination of static and dynamic risk factors included on a structured decision support tool (van den Berg et al., 2020; Dexter & Vitacco, 2020; Guy et al., 2015; Heffernan & Ward, 2015; Monahan & Skeem, 2016; Murrie & Kelley, 2017). The purpose and practice of contemporary violence risk assessment in many contexts hinges on the assumption that violence risk (state) is not only changeable but modifiable through interventions targeted at dynamic risk factors (Beech et al., 2016; Olver & Stockdale, 2020; Olver & Wong, 2011). Dynamic risk and protective factors possess great potential in their abilities to guide the selection of risk management strategies, identify treatment targets, monitor treatment progress, and monitor changes in risk over time (Cording et al., 2016; Klepfisz et al., 2016; McNaughton Nicholls & Webster, 2014).

However, the empirical literature addressing the key features of dynamic risk factors and risk state has fallen behind the theoretical and practice developments. In contrast to the vast number of studies that have examined the reliability and predictive validity of structured risk assessment tools, there are relatively few studies that have addressed whether dynamic risk factors on these tools change over time (i.e., internal responsiveness) or whether change on these dynamic risk factors results in corresponding change in violence (i.e., external responsiveness; Douglas & Skeem, 2005; Hanson & Harris, 2000; Monahan & Skeem, 2014; Olver & Stockdale, 2020; Serin et al., 2013; Skeem et al., 2000). Accordingly, the current research employed a longitudinal, multi-wave research design to address the internal, external, and comparative responsiveness of dynamic risk factors and ratings of risk state using the HCR-20 and START. Prior to discussing the findings and their implications, it is necessary to highlight several limitations of the current study.

4.1. Limitations and Qualifications

This section describes several limitations and qualifications that must be kept in mind while considering the relevance and strength of the current results and conclusions. The highlighted limitations and considerations include (a) the combination

of the two sub-samples, (b) the rate of attrition, (c) the nature of the assessments, (d) the risk assessment tools used, (e) statistical specifications and options, and (f) the nature of internal responsiveness.

4.1.1. Combination of Two Samples

The combination of data from two different types of samples must be noted. Ideally, sufficiently large samples sizes for both the civil psychiatric and correctional samples would have been available across all six assessments. This would have allowed for many of the analyses to be conducted separately within each of the two samples, as well as compared across the different sample types. Nevertheless, the samples were combined to provide a greater number of observations (particularly in the later follow-ups) for the analyses and resultantly increase the statistical power of the inferential procedures. This decision was also consistent with the nature of the current research questions.

It must be acknowledged, however, that meaningful and potentially relevant differences exist between civil psychiatric inpatients and general correctional offenders. For instance, in the current study, the correctional sample was generally rated as higher risk compared to the civil psychiatric sample. These differences may have an impact on the frequency, rate, or direction of change on dynamic risk ratings. On the other hand, there is considerable variability amongst both civil psychiatric patients and correctional offenders. This variability also likely impacts the frequency, rate, or direction of change on dynamic risk ratings. Potential differences in the within-group compared to between-group impacts are unknown. Additionally, many participants in the psychiatric sample reported a history of crime and criminal justice contact, and many participants in the correctional sample reported a history of mental illness and acute psychiatric treatment. Thus, in principle these participants may have been included in either sample, and this study happened to find them associated with a particular system (i.e., mental health or criminal justice). Although the combination of the two samples does not prevent conclusions from being drawn regarding the research questions of interest (i.e., internal and external responsiveness of dynamic ratings), this limitation must be kept in mind.

4.1.2. Attrition Rate

The rate of attrition and resultant missing data seen in the current study was another limitation. Due to attrition, the number of observations included in the analyses was considerably less than the full potential. Roughly three of four participants (74.5%) completed at least one follow-up assessment, while almost half (46.3%) completed all five follow-ups. The largest proportion of participants were lost after the baseline assessment with fewer lost during each of the follow-up intervals.

The attrition rate seen in the current study was comparable to the MacArthur study of mental disorder and violence (Monahan et al., 2001). Often thought of as a gold standard for longitudinal risk assessment research, a similar proportion of the sample (49.6%) completed all six assessments. The current attrition rate was also comparable to other longitudinal studies using the HCR-20. For instance, Penney and colleagues (2016) reported 75.9% of their sample completed the first follow-up after one-month post-discharge and 67.8% completed the second at six-months post-discharge. Neves and colleagues (2010) reported a slightly higher 85.4% of their sample completing a reassessment after six months. As well, Michel and colleagues (2013), using a research design in which participants could miss an assessment but complete the subsequent assessment, reported that 27.0% of the sample dropped out from the study, and they had 70.2% of the sample complete the final assessment at two-years post-discharge.

Although the attrition was comparable to other studies, the number of missing reassessments throughout the analyses was notable. With 216 participants completing the first assessment and six total assessments, there was the potential to examine change over 1080 pairs of assessments of which a total of 623 pairs of assessments were included in the current study (57.7%). It is possible that the rates of change reported in the current study might have been different if all participants completed all the assessments. More broadly, all reported results contain a degree of error related to missing data.

As discussed in the Statistical Analyses section, missing data for complete assessment periods was handled through pairwise deletion (RCI, Wilcoxon tests, Spearman's coefficients) or the use of procedures that can accommodate for different numbers of assessment across participants (GEE). Due to the nature of the research

questions and the nature of change analyses, multiple imputation was not utilized. Although the missing data limited the opportunities to observe change, the results can still provide some insights into the responsiveness of these dynamic ratings.

4.1.3. Nature of the Assessments

Another potential limitation of the study is the nature of the assessments. Although a rigorous research design (i.e., truly prospective, multi-wave) was used in the current study, the assessments were conducted by research assistants for research purposes only. The assessments did not impact the participants' release, management, treatment, or supervision in any way. This limitation is common. Many risk assessment studies are based on assessments conducted for research only, including several previous studies examining change on the HCR-20 and START (e.g., Coupland & Olver, 2018; de Vries Robbé et al., 2015; Hogan & Olver, 2016, 2019; Michel et al., 2013; Neves et al., 2010).

Additionally, the raters were students with varying levels of experience from senior undergraduate students to senior doctoral students. They all received training in general assessment principles and the use of the risk assessment instruments from one of the authors of each tool. Comparatively, some risk assessment studies have examined real-life or in-vivo assessments conducted by clinicians who are responsible for the management of the participants (e.g., Belfrage & Douglas, 2002; Belfrage et al., 2004; Draycott et al., 2012; Morrissey et al., 2014).

4.1.4. Risk Assessment Tools

The risk assessment tools, or decision support aids, focused on in the current study were the HCR-20 and the START. Specifically, version two of the HCR-20 was used, as version three was still under development and not implemented at the time data collection began (Douglas, Hart, et al., 2013). There are substantial differences between the versions of this tool and notable advances were implemented in the newest version. Although the risk factors were revised and updated, in general the same dynamic risk factors are included on version three. As well, studies have noted high correlations between scale and total scores across versions (Bjørkly et al., 2014; Douglas & Belfrage, 2014; Douglas & Shaffer, 2020).

The current study focused exclusively on the presence of dynamic risk and protective factors, as these are the only item level ratings made using version two of the HCR-20 or the START. However, the presence of a risk factor is not the only consideration. Version three of the HCR-20 also includes relevance ratings for each risk factor (Douglas, Hart, et al., 2013). Current theory postulates that risk factors may be present in a given individual, yet not relevant for that individual's risk for violence. Ultimately, it is those risk factors that are present and relevant which contribute to a given individual's risk state. Identifying which risk factors are relevant in each case allows for an understanding of what might have led the individual to perpetrate violence in the past and the development of plans to prevent future violence (Klepfisz et al., 2016). Determining the relevance of each dynamic risk factor also provides potential greater clarity on whether intra-individual change has occurred on risk factors that are functionally related to violence in that individual. For instance, Hogan and Olver (2019) found that change on dynamic risk factor relevance ratings were predictive of violence controlling for pretreatment total scores, whereas change on presence ratings were not predictive of violence when controlling for pretreatment total scores. As well, Hogan and Olver (2016) reported that change on both presence and relevance ratings were predictive of inpatient aggression while controlling for length of stay and pretreatment scores. However, the use of version two of the HCR-20 in the current study did not allow for the examination of relevance ratings.

Additionally, the overall process of completing a comprehensive risk assessment using the HCR-20 has experienced substantial development and direction with the new version. More guidance is provided in the overall process of completing the assessment, beginning with rating the presence and relevance of each risk factor, then developing an understanding of the case through formulation and scenario planning, linking to risk management strategies, and arriving at final opinions (Douglas, Hart, et al., 2013; Guy et al., 2015). On the other hand, in the current study using version two of the HCR-20, raters were only required to rate the presence of each risk factor and record a final opinion in the form of a summary risk judgment (i.e., Low, Moderate, or High Risk). That is, raters were not required to systematically record any details addressing case formulation, scenario planning, or management recommendations.

4.1.5. Statistical and Measurement Specifications

Several considerations should be noted regarding the statistical analyses, as well as some overarching considerations regarding the measurement of change. First, the measurement of the dynamic risk and protective factors using the noted three-point scale and the summation of these item scores to create scale scores is worth mention. These item ratings are notably made on an ordinal scale (i.e., No/Absent, Possible/Partial, Yes/Present) that is often coded numerically (i.e., 0, 1, 2). As mentioned, the individual risk factors or items are not meant to be scored numerically, nor are they intended to be summed into scale or total scores in clinical practice; however, numeric scores are typically used in research contexts to evaluate psychometric properties (Douglas, Hart, et al., 2013; Douglas et al., 2014).

Second, change scores were created from the raw score item ratings by subtracting the latter rating from the prior rating. As such, for the risk factors positive scores indicate improvement (i.e., decrease in presence), while negative scores indicate worsening (i.e., increase in presence). The creation of change scores in this manner ignores relative differences in risk. For instance, a change score of zero corresponds to individuals that did not change across assessments, with three ways to demonstrate no change (i.e., risk factor is Absent at both assessments, Possible at both assessments, and Present at both assessments). Similarly, a change score of one (or negative one) corresponds with individuals who increased (or decreased) in the presence of the risk factor by one-point, which meant there were two ways to arrive at a change score of one (i.e., risk factor is Present then Possible, or Possible then Absent). As such, the consideration of change scores potentially disregards important difference in relative standing on the risk factor, as well as equates the differences between positions on the three-point ordinal scale. It may be the case that certain changes are more important than others in terms of associated changes in the likelihood of violence. For example, a change from Absent to Possible could be more meaningful than a change from Possible to Present.

This also raises an important theoretical or philosophical question; is change more important than current standing on dynamic risk rating? Put another way, is there a difference in risk between two individuals both currently rated High risk, one of whom was previously rated High risk (no change across assessments) and the other was

previous rated Moderate Risk (increased in risk across assessments)? Focusing on change scores would mean that the latter individual, who increased in risk, would have a worse score and should subsequently be more likely to perpetrate violence. Statistically, the use of change scores to predict violence, therefore, treats these individuals as different and (partially) ignores relative standing on the most recent assessment. This complicates investigations into one of the defining criteria for a dynamic risk factor; change on a dynamic risk factor should be associated with corresponding change in the likelihood of violence (Andrews & Bonta, 2010b; Douglas & Skeem, 2005).

Third, interrater reliability was used in place of the test-retest reliability for several analyses addressing internal responsiveness, including the calculation of the SEM, MDC, and RCI. Interrater reliability was used due to its importance in the context of risk assessments. However, the interrater reliability for the current study was based on a small subset of only the civil psychiatric sample. As well, the standard deviations from all available assessments were used in these analyses, in contrast to using the standard deviation from the baseline (or first) assessments in each set of analyses.

With respect to the reliable change index analyses, the reliability was sufficiently high and standard deviation sufficiently low to result in two-point change on an item corresponding with reliable change for nearly all dynamic risk and protective factors on both tools. For one item on the HCR-20 and one item on the START, the combination of the relatively low reliability and relatively high standard deviation made it impossible for this item to detect reliable change. In contrast, for one item on the START, the combination of the relatively high reliability and relatively low standard deviation made it that any change on this item corresponded with reliable change. Accordingly, higher interrater reliability would ultimately lead to better performance in reliable change analyses. In this context, a high reliability combined with a low standard deviation would lead to the detection of partial change on all dynamic risk factors as reliable change. The nature of the three-point rating scheme for the items on these risk assessment tools may limit the utility of reliable change analyses on these ratings.

Fourth, GEE was used to address the external responsiveness of the dynamic ratings. As with similar generalized linear modeling techniques, this procedure examines the relationship between predictor variables and a dependent variable across multiple assessments. In this instance, the analyses were restricted to examining the relationship

between the dynamic risk ratings and violence reported at the next assessment (i.e., violence was lagged in the analyses, so baseline ratings predict violence at the first follow-up, ratings at the first follow-up predict violence at the second follow-up, and so on). For the analyses involving change scores, the analyses examined change from one assessment to the next predicting violence at the subsequent assessment (i.e., change between baseline and the first follow-up predicting violence at the second follow-up, and so on). As a result, the relationship between repeated assessments or change scores on the dynamic ratings and violence is dependent upon the length of the reassessment interval, which varied considerably across participants. Alternative models could have, for instance, examined the relationship with any future violence (i.e., violence reported at any subsequent follow-up assessment, not just the immediately succeeding follow-up).

Although the most direct examination of external responsiveness in this context entails an analysis of the relationship between change scores on a dynamic risk factor and violence, the utility of change scores may be limited. Ratings on these dynamic risk factors are ordered categorical. Although it is common to use numerical scores for research purposes (Douglas et al., 2014), calculating changes based on subtraction treats these ordered categorical ratings as if they were measured on an interval scale. Moreover, the use of change scores partially ignores the original ordering of the scores and potentially important differences in relative risk. On the one hand, change scores provide information regarding the presence of change (i.e., zero or non-zero value), direction of change (e.g., in this study positive scores indicate a decrease in the presence of a risk factor and negative scores an increase) and magnitude of change (one-point or two-point of change). On the other hand, change scores group unlike individuals (e.g., there were three ways to demonstrate no change across assessments) and differentiate between like individuals (e.g., two individuals with the same rating on the most recent assessment may have different change scores). Additionally, the use of change scores is artificial in the sense that it diverges from the typical use of the instrument in practice.

Finally, the current study involved many inferential tests within each section to address the various research questions without taking additional steps to control for errors across this number of tests. Each of the internal responsiveness analyses (i.e., Wilcoxon tests and Spearman's stability coefficients) involved 143 tests for the HCR-20 (13 variables across 11 assessment intervals) and 473 tests for the START (43 variables

across 11 assessment intervals). The external responsiveness analyses (i.e., GEE) involved 104 tests for the HCR-20 (four models each for 13 predictors across two dependent variables) and 344 tests for the START (four models each for 43 predictors across two dependent variables). Using a simple Bonferroni correction restricting familywise error to .05 would result in a per-test error rate of between roughly .0003 and .0001 for tests involving the HCR-20 and START, respectively.[16]

4.1.6. Nature of Internal Responsiveness

Internal responsiveness is conditional upon a number of methodological stipulations, including the specific procedures and measures employed in the study (Husted et al., 2000). A notable methodological factor with the potential for great impact on the internal responsiveness reported in each study is the amount of time between each of the assessments. The amount of change reported is inherently linked to the length of the reassessment interval.

In the context of ratings on a structured professional judgment tool, change is dependent upon the time frame of the assessments (i.e., the relevant period upon which the ratings are based) and the length of the reassessment interval. The length of the relevant time frame for the dynamic ratings limits the speed at which ratings can change. If the relevant time frame of the dynamic ratings is relatively long (e.g., rating the items based on the previous 12 months compared to previous month), ratings will not have the same opportunities to fluctuate. As well, some change will be seen while some change will be missed based on the length of the interval between assessments. Assessments spaced far apart may miss frequent oscillations, while assessments spaced closer together may not detect more gradual or prolonged change. Consequently, interpretations and generalizations of the results of a study reporting on internal responsiveness should be restricted to similar rating timeframes.

In the current study, assessments were intended to be spaced at one-month intervals, but the reassessment intervals were generally somewhat longer (six weeks). There was no maximum amount of time between reassessment intervals, and

[16] Note, the Bonferroni correction and resultant per-test error rates are included for comparative purposes and were not used in presenting or interpreting the current findings. As mentioned, the significance criterion was set at 0.05 for all hypothesis tests.

participants could not miss an assessment and then participate in a later assessment. Therefore, no matter the length of time that passed, each follow-up was recorded sequentially. As seen in Table 3-1, the average number of days across assessments ranged from 38 to 52 days with noticeable variability amongst participants (*SD* = 12 to 35 days). Comparisons from the baseline to each of the follow-up assessments allowed for evaluations across longer reassessment intervals, increasing to an average interval of 211 days (*SD* = 56).

4.1.7. Changes to Data Collection and Measurement

As the current study used data from two larger longitudinal projects, the data collection procedures were fixed prior to the origin of this study. If possible, several alterations would have been made to the procedures. Under ideal circumstances, changes would have been made to directly address some of the limitations described, including incorporating version three of the HCR-20 and attempting to decrease the attrition rate. However, many of these limitations were not directly addressable during the data collection process (i.e., version three of the HCR-20 was incorporated into the data collection as soon as the first draft was available,[17] and many strategies were already taken to reduce the attrition rate). The number of cases coded for interrater reliability would have been increased and additional details would have been captured regarding case formulation and scenario planning.

Additional changes would have been considered to allow for more in-depth analyses addressing internal and external responsiveness. One proposed change would have been to manipulate the reassessment interval and coding timeframe across participants. For instance, one possibility could have been to continue the data collection unchanged for half the sample with six assessments spaced approximately monthly. For the other half of the sample, data could have been collected using shorter reassessment intervals, such as 22 assessments approximately weekly. This could have allowed for additional examinations addressing speed of change and oscillations across a roughly equal five-month time interval.

[17] Specifically, version three of the HCR-20 was incorporated into the data collection several years after the beginning of the longitudinal projects; however, this data was not analyzed in the current study due to the limited sample sizes available across all assessments.

4.2. Dynamic Risk Factors on Structured Assessment Tools

The contemporary definition of a dynamic risk factor specifies four necessary criteria. A dynamic risk factor (a) precedes the outcome (e.g., violence), (b) increases the likelihood of the outcome, (c) is capable of change either spontaneously or because of intervention, and (d) demonstrates a relationship between change in the risk factor and the likelihood of the outcome (Andrews & Bonta, 2010b; Douglas & Skeem, 2005; Klepfisz et al., 2016). Accordingly, a dynamic risk factor must be capable of change, and change on the factor must correspond with change in the likelihood of future violence (Brown et al., 2009; Caudy et al., 2013; Douglas & Skeem, 2005; Hanson & Harris, 2000; Howard & Dixon, 2013; Jones et al., 2010; Lewis et al., 2013; Serin et al., 2016).

In conducting the current study, we aimed to add to the growing body of empirical research examining change on putatively dynamic risk factors included on structured violence risk assessment tools. The first research question addressed the internal responsiveness of dynamic risk factors. That is, this research question addressed the third criterion of a dynamic risk (or protective) factor: the ability to change over time either spontaneously or due to intervention (Andrews & Bonta, 2010b; Douglas & Skeem, 2005). Overall, the current results provide support for the internal responsiveness of many dynamic risk and protective factors on the HCR-20 and START. The second research question addressed the external responsiveness of dynamic risk factors. This question addressed the fourth criterion of a dynamic risk (or protective) factor: an association between change on the dynamic rating and change in the likelihood of violence (Andrews & Bonta, 2010b; Douglas & Skeem, 2005). Based on the current results, we found mixed or partial support for the external responsiveness of the dynamic factors on the HCR-20 and START.

4.2.1. Internal responsiveness of the HCR-20.

Generally, the current results add to the growing body of research supporting the internal responsiveness of the dynamic risk factors on the HCR-20. Although it was more common for a participant's rating on a given dynamic risk factor to remain constant from one assessment to the next, a considerable proportion of participants' dynamic risk factor ratings changed across assessments. Change from one assessment to the next over an average interval of 46 days was seen in 25% to 35% of reassessments for the

risk factors on the Clinical scale and 28% to 40% of reassessments for the risk factors on the Risk Management Scale. Michel and colleagues (2013) reported a comparable rate of change using considerably longer six-month reassessment intervals. They reported that items on the Clinical scale changed in 17% to 34% of reassessments in the forensic psychiatric sample and 20% to 44% of reassessments in the general psychiatric sample, while items on the Risk Management scale changed in 25% to 38% of reassessments in the forensic and 30% to 78% of reassessments in the general psychiatric sample. Markedly, the current study reported higher proportions of change for some items compared to Michel et al. (2013) even though the current reassessment interval was considerably shorter.

This study also found that it was common for participants' ratings to change at least once on the dynamic risk factors across all six assessments. That is, 51% to 75% of participants' ratings on the Clinical items and 58% to 77% of participants' ratings on the Risk Management items changed at least once. Many participants changed once (24% to 37%) or twice (11% to 29%) across the five possible reassessments on each dynamic risk factor. Notably, a minority of participants' ratings changed repeatedly across the assessments, with some participants changing in four (3% to 8%) or all five (0% to 2%) possible instances. Michel and colleagues (2013) reported similar results across five assessments spaced six months apart. Depending on the sample and item, they found that 42% to 83% of participants scores on the Clinical items changed at least once and 58% to 81% of participants scores on the Risk Management items changed at least once. Although the analyses and frequency of change reported varied across studies, both the current study and Michel et al. (2013) reported that participants were least likely to change on the same item (Negative Attitudes). This item consistently yielded the lowest frequencies of change from one assessment to the next and across all assessments.

Overall, substantial change has been observed based on the proportions of reassessments demonstrating change in dynamic risk factors. A minority of the current sample (17% to 47%) changed on the dynamic items across every assessment interval. Generally, it was much more common for participants to experience one-point change (17% to 45% of reassessments) than two-point change (0% to 8% of reassessments) on the item ratings. That is, changing from a risk factor being Absent to Present or vice versa was relatively rare compared to partial changes in the presence of the risk factor.

In contrast to raw score change, reliable change on the dynamic risk factors was relatively rare in the current study. Except for a single item, reliable change on the dynamic risk factors corresponded with full or two-point change on the item rating. This means the risk factor must change from Absent to Present or vice versa to qualify as reliable change. Although reliable change was infrequent, the results still support the internal responsiveness of the dynamic risk factors on this tool. Reliable change was seen in a small proportion of participants (1% to 7.5%) for all but one item across nearly every reassessment interval. Across all assessment intervals, reliable change occurred in 0.9% to 4.5% of the dynamic risk factors ratings. Importantly, one item (Active Symptoms of Major Mental Illness) was not able to detect reliable change. No identified studies have previously reported results of reliable change analyses on the individual dynamic risk factors.

Internal responsiveness was also investigated using Wilcoxon signed-rank tests. Based on these analyses, three items on the Clinical scale (Negative Attitudes, Active Symptoms of Major Mental Illness, and Impulsivity) and three items on the Risk Management scale (Exposure to Destabilizers, Noncompliance with Remediation Attempts, and Stress) demonstrated change across assessments. For some of these items, change was only evident during certain reassessment intervals. Three items (Active Symptoms of Major Mental Illness, Impulsivity, and Noncompliance with Remediation Attempts) evidenced the most consistent mean rank change.

Comparing the aggregate tests to the proportions, it is apparent that considerable intraindividual change was not detected using the mean rank tests. Although only three risk factors on each of the HCR-20 scales were found to change using the Wilcoxon tests, all items demonstrated roughly comparable frequencies of raw score and reliable change. For instance, Negative Attitudes changed in fewer reassessments (25%) than any other risk factor when considering all possible assessments, yet this risk factor demonstrated mean rank change. One the other hand, Unresponsive to Treatment changed in 35% of reassessments and displayed some of the largest rates of reliable change, yet it did not demonstrate mean rank change.

Generally, the prior research used longer reassessment intervals and comparable statistical techniques. Six previous studies have investigated aggregate changes on HCR-20 risk factors, using Wilcoxon signed rank tests, Friedman's tests, or

linear mixed-modeling. One study did not find any change. Belfrage et al. (2004) reported no change on any dynamic risk factors on the HCR-20 across two assessments over a three to 24-month period (*M* = 12 months). Although this study did not find any change in risk factor ratings over time, the remaining five also found support for the internal responsiveness of some of the dynamic risk factors.

Neves and colleagues (2010) examined change across three assessments at roughly six-month intervals. They found change in only one Risk Management item (Stress) over time. In contrast, Olsson and colleagues (2013) reported change in all ten dynamic items from the first to the most recent assessment over an average of 43 months (range = 2 to 146). However, when the Risk Management items were rated for an institutional setting, only three items on this scale (Plans Lack Feasibility, Exposure to Destabilizers, and Noncompliance with Remediation Attempts) changed from the first to most recent assessment. As well, when examining change from the first to the subsequent assessment, four items on the Clinical (all except Impulsivity) and four items on the Risk Management scale (all except Lack of Personal Support) demonstrated mean rank change across the average nine-month interval.

Belfrage and Douglas (2002) reported change in fewer items on the HCR-20 across three assessments at six-month intervals. When examining change from the first to the second assessment, they found no mean change on any of the Clinical or Risk Management items when rated for a community setting, but mean change was seen on two Risk Management items (Exposure to Destabilizers and Noncompliance with Remediation Attempts) when rated for an institutional setting. In contrast, when examining change from the first to the third assessment, they found mean change on two Clinical items (Negative Attitudes and Unresponsive to Treatment) and no Risk Management items when rated for a community setting, but mean change was seen on one Risk Management item (Plans Lack Feasibility) when rated for an institutional setting. In addition, Michel and colleagues (2013) investigated change across five assessments using Friedman tests. They observed change on two Clinical and four Risk Management items, with the items varying across the two samples. O'Shea and Dickens (2015) examined change across up to four assessments with an average length of 182 days (*SD* = 48) between assessments. Using linear mixed effects modeling, they reported mean change in four Clinical items (all except Unresponsive to Treatment), but none of the Risk Management items.

Largely, the available empirical evidence supports the internal responsiveness of the dynamic risk factors on the HCR-20. Research is accumulating that observed change on these items over time, which satisfies this necessary or defining feature of a dynamic risk factor. In addition, as these items have been shown to change over time, they would be appropriately labelled as variable risk markers or variable risk factors according to Kraemer and colleague's (1997) typology. This distinction is based on the ability to change as a direct result of intervention (variable risk factor) or only change spontaneously (variable risk marker). Although much of the available research included individuals undergoing some form of supervision, treatment, or management (e.g., correctional offenders on probation, or forensic psychiatric patients discharged to community under supervision), the specific nature and targets of the interventions and management strategies are largely unknown. As such, the term variable risk marker would be appropriate, as the underlying reasons for change are unknown.

4.2.2. External responsiveness of the HCR-20.

Overall, the current results found partial support for the external responsiveness of several dynamic risk factors on the HCR-20 and add to the limited research that has examined this aspect of responsiveness. Broadly, some support was found for the external responsiveness of four Clinical items (Lack of Insight, Active Symptoms of Major Mental Illness, Impulsivity, and Unresponsive to Treatment) and all Risk Management items. However, support for the external responsiveness of these factors was only seen in some analyses and the most direct examination of external responsiveness identified few items.

Ultimately, external responsiveness concerns the extent to which change on a dynamic risk factor is associated with change in the likelihood of violence. The most direct examination of the relationship between change on each of the risk factors and violence in separate bivariate GEE models failed to support the external responsiveness of most items. Change scores on one item (Active Symptoms of Major Mental Illness) were predictive of overall violence across assessments and change scores on two items (Lack of Insight, and Active Symptoms of Major Mental Illness) were predictive of serious violence across assessments. The relationship between change on these specific dynamic risk factors and violence, or their external responsiveness, was directly evidenced in these analyses.

External responsiveness was also (indirectly) examined by analyzing the relationship between repeated assessments on the dynamic risk factors and violence while controlling for the baseline rating on the respective risk factor in separate GEE models. These analyses provided support for the external responsiveness of several dynamic risk factors on the HCR-20. When controlling for baseline ratings, three Clinical risk factors (Lack of Insight, Impulsivity, and Unresponsive to Treatment) and all Risk Management factors were predictive of overall violence across assessments. Two Clinical (Lack of Insight and Impulsivity) and two Risk Management (Plans Lack Feasibility and Exposure to Destabilizers) items were predictive of serious violence across assessments when controlling for their respective baseline ratings. In other words, a participant's current rating on each of these risk factors was predictive of violence above and beyond their original (baseline) rating on that risk factor. The fact that the most recent or current standing on a given risk factor added incrementally to the baseline rating for this risk factor implies that there was some difference between the baseline and most recent ratings. However, these analyses do not directly address the relationship between change on the item and violence.

While a growing body of research is accumulating regarding the external responsiveness of the HCR-20 generally, only a single study was identified that reported on the individual dynamic risk factors. Michel and colleagues (2013) used a similar approach to assess the external responsiveness of items on the HCR-20. They examined the relationship between change scores on each of the dynamic risk factors and violence across five assessments in separate bivariate GEE models. They observed that change scores on three Clinical factors (Negative Attitudes, Impulsivity, and Unresponsive to Treatment) and three Risk Management factors (Plans Lack Feasibility, Lack of Personal Support, and Noncompliance with Remediation Attempts) were predictive of violence.

Largely, the available empirical evidence provides limited support for the external responsiveness of some of the dynamic risk factors on the HCR-20. Few studies have directly examined the relationship between change on the individual items and violence. Nevertheless, the available research provides support for the external responsiveness of some dynamic risk factors in certain analyses. Accordingly, empirical evidence is beginning to accrue that supports this final defining feature of a dynamic risk factor. In addition, it may be appropriate to label these items as causal risk factors according to

Kraemer and colleague's (1997) typology. However, the available empirical evidence supporting the final requirement for a dynamic risk factor or causal risk factor using the HCR-20 is limited.

4.2.3. Internal responsiveness of the START.

Overall, the current results support the internal responsiveness of the dynamic risk and protective factors on the START. Few studies have previously investigated item level change on the START. Only one prior study using the adolescent version of the START was identified that reported change in the individual item ratings over time (Sellers et al., 2017). As such, the current study is the first known to report on the internal responsiveness of the individual items on the adult version of the START.

Like the HCR-20, it was generally more common for a participant's rating on a given dynamic risk or protective factor on the START to remain constant from one assessment to the next. Nevertheless, a notable proportion of participants' item ratings changed across assessments. Considering all possible reassessments, change was observed in 24% to 42% of reassessments for the risk factors on the Vulnerability scale and 20% to 30% of reassessments for the protective factors on the Strength scale. It was also common for participants' ratings to change at least once on the dynamic factors across all six assessments. Namely, 51% to 76% of participants scores on the Vulnerability items and 60% to 74% of participants scores on the Strength items changed at least once. On each item, most participants changed once (21% to 34%) or twice (14% to 30%) across the five possible reassessments. Once again, a minority of participants' ratings changed repeatedly across the assessments, with some participants changing in four (2% to 11%) or all five (0% to 3%) possible instances. No identified studies have previously reported the frequency of raw score change on the Vulnerability and Strength factors.

Substantial change was observed based on the proportions of reassessments demonstrating change in dynamic risk and protective factors on the START. A minority of the sample (17% to 45%) changed on each of the START dynamic items across every assessment interval. The same pattern was evident for the START as for the HCR-20. It was more common for participants to experience one-point change (16% to 47% of

reassessments) than two-point change (0% to 9% of reassessments) on the item ratings. That is, full change on an item was rare compared to partial change.

In contrast to raw score change, reliable change on the dynamic items was comparatively rare on the START. Except for two items on the Vulnerability scale, reliable change on the dynamic items corresponded with full or two-point change on the item rating (i.e., change from Absent to Present or vice versa). Akin to the results for the HCR-20, although reliable change was infrequent, the results still support the internal responsiveness of the dynamic factors on the START. Reliable change was seen in a small proportion of participants (1% to 8.3%) for all but one item across nearly every reassessment interval. Across all assessment intervals, reliable change occurred in 0.2% to 5.3% of item ratings. Importantly, one item on the Vulnerability scale (Mental State) was not able to detect reliable change. Even full change on this item did not qualify as reliable change. In contrast, one item on the Vulnerability scale (Substance Use) was sensitive enough such that all change corresponded with reliable change. None of the identified studies have previously reported results of reliable change analyses on the START items.

Additionally, based on the Wilcoxon signed-rank tests, change was seen across at least one assessment interval on 13 items on the Vulnerability scale and 15 items on the Strength scale. For most of these items, change was only seen from the baseline to the first follow-up assessment (13 Vulnerability and 12 Strength items). It was rare for an item to change across more than one of the possible five assessment intervals; only one Vulnerability and four Strength items changed twice. When compared from the baseline assessment, six Vulnerability and four Strength items demonstrated changed from baseline to each of the follow-up assessments.

The current study reported change on more dynamic factors than the only prior study to evaluate change on START items over time. Sellers and colleagues (2017) examined item level change on the adolescent version of the START over assessments spaced approximately three months apart (M = 115 days, SD = 78). They reported mean change on four Vulnerability items and six Strength items. Although few studies are available, the available empirical evidence supports the internal responsiveness of some of the dynamic risk and protective factors on the START. The current and prior study observed change on several of these items over time, which satisfies this defining

feature of a dynamic risk (or protective) factor. Like the HCR-20, as these items have been shown to change over time but the underlying reasons for this change are unknown, they would be appropriately labelled as variable risk markers according to Kraemer and colleague's (1997).

4.2.4. External responsiveness of the START.

Overall, the current results found partial support for the external responsiveness of some of the dynamic risk and protective factors on the START. Largely, some support was found for the external responsiveness of 14 Vulnerability and 12 Strength items. However, support for the external responsiveness of these factors was dependent upon the outcome of focus and the type of analysis. Once again, the most direct examination of external responsiveness identified few items. Change scores on one item (Self-Care on the Vulnerability scale) were predictive of overall violence across assessments and change scores on one item (Materials Resources on the Vulnerability scale) were predictive of serious violence across assessments.

Like the HCR-20, the external responsiveness for more items on the START was supported indirectly through comparisons of the current ratings controlling for baseline ratings. That is, when controlling for baseline ratings on the respective item, 12 Vulnerability and 10 Strength items were predictive of overall violence across assessments, while 12 Vulnerability and seven Strength items were predictive of serious violence across assessments. While a body of research is accruing regarding the external responsiveness of the START, none of the identified studies have reported on the individual dynamic risk factors.

The current empirical evidence provides limited support for the external responsiveness of some of the dynamic risk and protective factors on the START. Generally, the current study provided partial support for the external responsiveness of just over half of the items on this tool. As no other studies have examined the external responsiveness of items on the START, conclusions regarding the appropriate classification of the items on this tool are limited.

4.2.5. Speed of Change versus Proneness to Change

As seen in Table 1-1, theoretical distinctions have been made between dynamic risk factors based on their supposed speed of change. Stable dynamic risk factors change gradually over longer periods of time compared to acute dynamic risk factors, which are capable of rapid change over short periods of time (Hanson & Harris, 2000; van den Berg et al., 2020). As mentioned, although this distinction is often discussed in the literature, few studies have empirically evaluated differences in the speed of change between different dynamic risk factors (i.e., between stable and acute factors) and some authors have questioned this distinction altogether (Mann et al., 2010; Thornton, 2016).

Some conclusions can be drawn from the current study regarding the speed of change in dynamic risk factors on these instruments. No clear increase or decrease was seen in the proportion of participants who changed on any items as the length of the reassessment interval increased (see Tables 3-14 and 3-16). That is, the proportions of participants who experienced reliable change on the dynamic risk and protective factors were comparable when examined from the baseline to each of the follow-up assessments. Although the length of interval increased, there was no obvious patterns in the rates of reliable change. The rates of reliable change varied slightly across the assessment intervals, even though the reassessment interval increased from an average of 53 to 211 days. Based on these observations, there is little support for the categorization of these risk factors based on speed of change.

However, there are several issues with drawing conclusions about the speed of change on these items from the available data and analyses. Although the average reassessment interval increased from the baseline to each of the follow-up assessments, there was also considerable intraindividual and interindividual variation in the length of each of the reassessment intervals. For instance, some participants completed several follow-up assessments within the same number of days in which other participants completed a single follow-up assessment. As well, it is possible that the length of the reassessment intervals used in this study were too long to capture change on acute dynamic risk factors. Acute dynamic risk factors are thought to be capable of change quickly in a matter of weeks, days, or hours (Hanson & Harris, 2000; van den Berg et al., 2020). Theoretically, it may be the case that all the dynamic risk and

protective factors included on these tools are stable dynamic risk factors, these factors are thought to require effortful processing or targeting interventions to change.

Another important consideration in the context of dynamic ratings on a structured professional judgment instrument is the inherent relationship between the length of the reassessment interval and the length of the rating timeframe (i.e., the relevant length of time upon which the ratings are based). This consideration has implications regarding the operationalization of the dynamic risk factors on these instruments and the minimum rating timeframe that would be appropriate. An appropriate rating timeframe should be informative, meaningful, and predictive of violence. The use of too short a time frame may obfuscate the true nature of the risk factor, while too long of a timeframe may dilute the ability of the factor to demonstrate change.

The distinction between dynamic factors based on speed of change may be an oversimplification. There are several important characteristics of change that may vary between dynamic risk factors in addition to the speed of change. For instance, dynamic risk factors may vary with respect to their likelihood or proneness to change and could potentially be categorized based on the general proneness to change. Some risk factors may be more inclined to change compared to others regardless of the length of the reassessment interval. This potentially important characteristics of dynamic risk factors relates to their comparative responsiveness.

The third research question addressed the comparative responsiveness of select dynamic risk factor ratings. These analyses investigated potential differences in the frequency of change and reliable change seen across select risk factors on the HCR-20 and can speak to the proneness to change of these risk factors. Generally, two risk factors that were more prone to change (Unresponsive to Treatment and Stress) and two risk factors that were less prone to change (Negative Attitudes and Lack of Personal Support) were identified. Unresponsive to Treatment changed in more instances than two other items and experienced more reliable change than five other items. Stress experienced more change than six other items and more reliable change than three other items. In contrast, Negative Attitudes changed in fewer instances than all but one other risk factor, yet only experienced less reliable change than one other risk factor. Lack of Personal Support experienced less change than seven other items and less reliable change than four other items. These items also differed in the number of times

change was observed across the six assessments, supporting the same general pattern. As such, it may be important to consider general differences in the proneness to change across different dynamic factors.

4.2.6. The multifaceted nature of violence.

The focus of the current research was the individual risk factors, yet violence is multifaceted. A single risk factor is not expected to account for all violence or be relevant to violence in all individuals. For several decades, it has been understood that people will perpetrate violence as a result of the presence of diverse arrays of risk factors and "there is no single path in a person's life that leads to an act of violence" (Monahan et al., 2001, p. 142). It is well accepted that the "causes of aggression and violence comprise a number of interacting processes, none of which are necessary or sufficient for violence to occur" (Klepfisz et al., 2016, p. 132).

Klepfisz and colleagues (2016) argued "it is unrealistic to expect change in any single dynamic risk factor to predict recidivism," as "violence is related to an array of risk factors and we would need to see change in various areas for there to be meaningful reductions in the risk for violence" (p. 132). To fully understand the influence of a given risk factor, it must be understood in the context of the other risk factors (Kraemer et al., 2001). As a result, it is important to consider the relationships amongst the risk factors for violence, and the relationships amongst change on these risk factors over time (Kraemer, 2003; Kraemer et al., 1997, 2001).

One of the current research questions addressed the interrelationships amongst dynamic risk factors over time. Specific analyses examined the associations amongst ratings and change scores on the dynamic risk factors on the HCR-20 over time. With respect to the relationship between ratings over time, most dynamic risk factors were correlated across the six assessments with coefficients in the small range. Some of strongest interitem relationships were consistent with the understanding and operationalization of these factors. Unresponsive to Treatment and Noncompliance with Remediation Attempts displayed the strongest relationship, and both assess the same general characteristics over different time periods (current and future). Lack of Insight and Unresponsive to Treatment also displayed one of the strongest item level relationships. The impact of insight on treatment compliance and vice versa is

undeniable. Stress and Lack of Personal Support also displayed a relatively strong relationship. In addition, consistent with the tool and underlying theory, the strongest relationships were found between the scale scores and the respective items on that scale, as well as between the summary risk judgements and all the other ratings. Many of these same patterns were evident when examining the relationship between change scores on the HCR-20 dynamic items over time.

Kraemer and colleagues (2001) described three important characteristics of a set of risk factors, including temporal precedence, correlation, and dominance. These features are important in understanding and prioritizing dynamic risk factors (Douglas & Skeem, 2005). The nature of the data collection and analyses only allowed for the examination of one of these features. These results add to the evidence supporting the multifaceted nature of violence and the importance of considering risk factors in context. Individual risk factors do not often work in isolation but operate synergistically to bring about violence. Most of the risk factors on the HCR-20 were correlated over time, as was change on these items.

4.3. Risk State on Structured Risk Assessment Tools

The contemporary understanding of violence risk acknowledges that an individual's risk fluctuates over time and circumstance. As discussed, the distinction between risk status and risk state has been increasingly accepted (Douglas & Skeem, 2005; Heilbrun et al., 2009; Klepfisz et al., 2016; Monahan, & Skeem, 2016; Skeem & Mulvey, 2002). Risk state signifies the overall risk level posed by a particular individual at a given moment in time based on their current standing on a combination of static and dynamic risk factors. Accordingly, risk state varies over time and circumstance due to external (e.g., treatment, supervision, monitoring) and internal (e.g., learning, aging, motivation) forces.

The fundamental aim of a comprehensive violence risk assessment rests on the notion that risk state is capable of being altered by external forces. Assessing risk in order to prevent future violence necessitates being able to alter the individual's risk (i.e., using strategies targeted at dynamic risk factors to reduce their likelihood of violence). The fundamental aim of many policing, correctional, and forensic interventions also rest

on the same notion. These interventions aim to reduce the overall likelihood the individual will engage in violence or crime through targeted strategies.

The current study aimed to add to the limited body of empirical research examining change on global assessment of risk state on structured violence risk assessment tools. On the HCR-20 and START, risk state is represented by the summary risk judgments. The first research question addressed the internal responsiveness of risk state ratings. That is, this research question addressed the extent to which these ratings were able to change over time either spontaneously or due to intervention. Overall, the current study found support for the internal responsiveness of the summary risk judgments on the HCR-20 and START. The second research question addressed the external responsiveness of risk state ratings. This question addressed the extent to which change on the summary risk judgments was associated with change in violence. Overall, the current study found partial support for the external responsiveness of the summary risk judgments on the HCR-20 and START.

4.3.1. Internal responsiveness of risk state ratings.

Generally, the current results add to the growing body of research supporting the internal responsiveness of the summary risk judgments on the HCR-20 and add the first identified study to focus on these ratings on the START. Although it was more common for ratings on the summary risk judgments to remain constant from one assessment to the next, a considerable proportion of participants' ratings changed across assessments. Change from one assessment to the next was seen in 21% of reassessments for the summary risk judgments on the HCR-20 and 20% for the summary risk judgments on the START. It was common for the summary risk judgments to change at least once across all six assessments. That is, 45% of participants summary risk judgments on the HCR-20 changed at least once and 42% of participants ratings on the START changed at least once. A notable portion of participants changed once (23% and 20%) or twice (11% and 12%) across the five possible reassessments, and a minority changed repeatedly with some participants changing in four (2.5% and 0.6%) or all five (0.6% and 0.6%) possible instances on the HCR-20 and START, respectively.

It was much more common for participants to experience one-point change (12% to 27% of reassessments) than two-point change (0% to 3% of reassessments) on the

summary risk judgments. That is, changing from Low to High risk or vice versa was rare compared to partial changes (i.e.., Low to Moderate or Moderate to High or vice versa). Relatedly, and in contrast to raw score change, reliable change on the summary risk judgments was relatively rare. Like the dynamic risk factors, reliable change on the summary risk judgments corresponded with full or two-point change on the rating. Meaning the rating must change from Low to High or vice versa to qualify as reliable change. Although reliable change was infrequent, the results still support the internal responsiveness of the dynamic risk factors on this tool. Reliable change was seen in a small proportion of participants across most reassessment intervals. Internal responsiveness was also investigated using Wilcoxon signed-rank tests. Based on these analyses, the summary risk judgments on the HCR-20 and START changed from the baseline to each of the follow-up assessments. Change between follow-up assessments was only seen for one assessment interval on the START and none on the HCR-20.

The current study found slightly more change on these ratings that the three prior studies using the HCR-20. Neves and colleagues (2010) observed change in 15% of summary risk judgments across three assessments at roughly six-month intervals, Coupland and Olver (2018) observed change in 14% over two assessments six months apart, and O'Shea and Dickens (2015) observed change in 11% of summary risk judgments across two to four assessments with an average six-month reassessment interval. O'Shea and Dickens (2015) similarly reported that two-point change was rare (1%) compared to partial change (10%). Noticeably, the current study reported higher proportions of change compared to both these studies even though the current reassessment interval was considerably shorter. As well, Penney and colleagues (2016) reported an increase in the proportion of Low ratings (18% to 38%) and a decrease in the proportion of Moderate ratings (76% to 54%). On the other hand, the current study found less change on these ratings than the one prior study using the START. Viljoen and colleagues (2012) observed change in 40% of summary risk judgments using the adolescent version of the START over a three-month assessment interval.

Largely, the limited available empirical evidence supports the internal responsiveness of global assessment of risk state on the HCR-20 and START. Research is accumulating on the HCR-20 and START that reported change on the summary risk judgments over time. The ability to change over time is a defining feature of ratings of risk state. Risk state is thought to vary over time and context as the

individual experiences change in relevant risk domains. Accordingly, it would be appropriate to refer to the summary risk judgments made using these tools as representing risk state, as opposed to representing risk status.

4.3.2. External responsiveness of risk state ratings.

Overall, the current results provide partial support for the external responsiveness of the summary risk judgments on the HCR-20 and START. Few studies have previously investigated this feature of these ratings. Only one prior study has examined the relationship between change on the summary risk judgments and violence using the HCR-20 (Coupland & Olver, 2018). As such, the current study is the first known to report on the external responsiveness of the summary risk judgments on the START and adds to the limited available research on the HCR-20. Broadly, some support was found for the external responsiveness of the summary risk judgments on both tools. However, support for the external responsiveness of these ratings was only seen in some analyses and the most direct examination did not support their external responsiveness.

The most direct examination of external responsiveness assessed the relationship between change scores on the summary risk judgments and violence in separate bivariate GEE models. Change scores based on the summary risk judgments were not predictive of violence across assessments. The relationship between repeated assessment of risk state and violence was also examined in other models. The results for the summary risk judgments made using the HCR-20 and START were alike. The summary risk judgments were predictive of both definitions of violence across assessments in separate bivariate analyses. The current standing on the summary risk judgments was also predictive of both definitions of violence when controlling for baseline summary risk judgment. In other words, an individual's current rating of risk state was predictive of violence above and beyond their original (baseline) rating of risk state. The current standing on the summary risk judgments was also predictive of both definitions of violence when controlling for baseline Historical scale scores. The Historical scale provides an estimate of risk based on static risk factors corresponding to risk status. As such, the current risk state estimate was predictive of violence when controlling for risk status.

While research is accumulating regarding the external responsiveness of the HCR-20, only a single study was identified that reported on the summary risk judgments. Coupland and Olver (2018) examined the relationship between change in summary risk judgments and various outcomes. Using a pseudo-prospective design, they examined change over two assessments six months apart. The results were dependent upon the outcome of interest. For the community outcomes, change in summary risk judgments was predictive of general recidivism, but not violence. For the institutional outcomes, change in summary risk judgments was predictive of serious institutional misconduct, but not any institutional misconduct. No identified studies have reported on this aspect of the summary risk judgments using the START.

Largely, the available empirical evidence provides partial support for the external responsiveness of global assessment of risk state made using the HCR-20 or START. Few studies have directly examined the relationship between change on the summary risk judgments and violence, or between repeated assessments on the summary risk judgments and violence. Nevertheless, the available research provides some support for that notion that fluctuations in ratings of risk state are associated with violence, and that current risk state is associated with violence above and beyond prior ratings of risk state or ratings of risk status. This empirical evidence also supports the distinction between risk status and risk state, as well as the defining features of risk state as assessed using summary risk judgments on the HCR-20 and START.

4.4. Implications for Risk Assessment and Management

Current practice emphasizes the dynamic nature of violence risk and many underlying risk factors for violence (Doyle & Logan, 2012; Douglas & Skeem, 2005; Serin et al., 2016; Skeem et al., 2000; Webster et al., 2000). It is common practice to incorporate dynamic risk factors to guide case formulation and treatment selection (Heffernan & Ward, 2015, Ward & Beech, 2015). Many contexts necessitate a comprehensive violence risk assessments stressing the short-term prediction of violence, identification of treatment targets, and implementation of management strategies to reduce risk (Andrews & Dowden, 2007; Doyle & Logan, 2012; Hanson & Harris, 2000). A focus on dynamic ratings is essential to implement risk management strategies, monitor progress, and ultimately, prevent violence (Dexter & Vitacco, 2020; Douglas & Skeem, 2005; Guy et al., 2015; Monahan & Skeem, 2016; Ward, 2016).

There are few guidelines offered to professionals regarding various aspects of dynamic risk (Douglas & Skeem, 2005). Some structured instruments include theoretical rational for the appropriate assessment timeframe for dynamic ratings and the appropriate reassessment interval (Douglas, Hart, et al., 2013). However, the ultimate decision often rests with the professional or can be dictated by legal or policy requirements. Douglas and Reeves (2010) describe some general guidelines regarding the reassessment of risk using the HCR-20. It is recommended that high risk individuals be reassessed more frequently than low risk individuals. As well, a new assessment should be conducted whenever any significant life changes occur.

Empirical research examining change in violence risk is complex (Olver & Stockdale, 2020). Empirical results regarding the proportion of individuals who changed or aggregate level change on a dynamic rating over a specific period must be incorporated into practice. The available empirical research on the HCR-20 and START has used reassessment intervals ranging from just over a month to several years. Change has been observed on some dynamic ratings over weeks, while no change has been observed on the same ratings over years (e.g., Holliday et al., 2012; Draycott et al., 2012). Additional longitudinal research is needed in this area to better understanding change on these dynamic risk factors over time. Dynamic risk factors may differ in their changeability, including the speed, likelihood, and magnitude of change. Change on a single dynamic risk factor is likely dependent upon several other dynamic risk factors, and change is likely impacted by various internal and external forces.

The importance of ongoing assessments monitoring dynamic risk factors to provide reactive and targeted interventions has been acknowledged for several decades (Hanson & Harris, 2000). The available research provides support for the use of the HCR-20 and START to monitor change in dynamic risk factors and risk state over time. As mentioned, it may be appropriate to label some of the risk factors on the HCR-20 and START as meeting all the criteria for a dynamic risk factor (Douglas & Skeem, 2005) and the criteria for a causal risk factor (Kraemer et al., 1997). Causal risk factors are of paramount importance, as ultimately professionals want to provide management strategies to reduce or mitigate these risk factors to lessen the individual's risk state.

A better understanding of change in risk state and dynamic risk factors will provide additional guidance regarding monitoring intervention progress, altering

intervention targets, implementing new management strategies, and reducing existing management strategies. Selecting an appropriate and defensible assessment schedule for a given individual requires the consideration of several factors (Douglas & Skeem, 2005). Professionals need to consider the potential speed and likelihood of change in dynamic risk factors and risk state. This consideration will be dependent in part on determining a meaningful assessment timeframe and reassessment schedule. Importantly, the forces that may be driving change must be understood. Both characteristics of the individual and characteristics of the risk assessment tool will impact the appropriate assessment schedule.

Olver and Stockdale (2020) highlighted the potential benefits of incorporating empirically supported dynamic violence risk assessment tools into standard practice; "to aid case formulation, to monitor progress toward risk reduction; and to enhance the utility, accuracy, fairness, and safety of release and reintegration decisions" (p. 7). Comprehensive violence risk assessments necessitate the consideration of dynamic risk factors and risk state (Andrews & Bonta, 2010a, 2010b; Craig et al., 2005; Serin et al., 2016; Wilson et al., 2013). Repeated assessments of dynamic risk factors and risk state allow for targeted interventions to be selected, the effectiveness of these interventions to be monitored, and the intensity or target of these interventions to be altered, all based on individualized management plans (Bonta, 2002).

Comprehensive assessments using a structured professional judgment instrument focus on more than simple lists of dynamic risk factors. Case formulation serves as the link between the presence of individual dynamic risk factors, the selection of targeted management strategies, and the overall understanding of risk state (Bonta, 2002; Serin et al., 2016). As Mann and colleagues (2010) stressed, "assessments respond better to the needs of decision makers and those being assessed (and to science) when the evaluation also explains the source of the risk" (p. 192).

Evaluation of the presence of a set of dynamic risk (and protective) factors is a limited portion of a comprehensive assessment. Dynamic risk factors must be understood idiographically in relation to violent behaviour, and violence must be understood in terms of its underlying goals or aims (Heffernan & Ward, 2015). Determining the relevance of each dynamic risk factor is a critical step in developing hypotheses about what lead this specific individual to perpetrate violence in the past and

how to prevent them from perpetrating violence in the future (Guy et al., 2015; Klepfisz et al., 2016). Relevance ratings identify those variables that are functionally related to violence. Relevance ratings begin to move the assessment from a descriptive account of risk factors to more explanative account of the case.

Some authors have stressed that the current operationalization of some dynamic risk factors refer to broad terms that consists of qualitatively different elements each associated with distinctive causal pathways (Ward, 2015, 2016; Ward & Beech, 2015). Although these authors argue that this limits their explanative power, broad conceptualizations allow for the identification of specific manifestations of each risk factor with a given individual. These specific manifestations are then considered in formulating an overall understanding of the case. Relevance ratings and case formulation require that each risk factor be broken down into its components to determine their functional relation to violence. Through formulation, professionals can "break down dynamic risk factors into several causal elements that in certain environments create, and maintain, antisocial values or behavior" (Ward, 2015, p. 109). Case formulation takes a list of dynamic risk factors and provides an explanative account of the relevant information. In the context of ongoing or repeated assessments, formulation and relevance ratings allows for greater potential in monitoring and understanding change in risk.

4.5. Conclusions and Future Directions

Professionals have increasingly called attention to the need for additional research investigating the changeability of dynamic risk factors and global estimates of risk state over time (Douglas & Skeem, 2005; Olver & Stockdale, 2020; Serin et al., 2016). The amount of theoretical literature discussing dynamic risk still dwarfs the empirical literature that truly examines dynamic risk. The demand for risk assessments aimed at monitoring and managing violence are increasing and the science must keep up with practice (Douglas & Kropp, 2002). A foremost task facing the field is the evaluation of how best to evaluate dynamic risk factors and monitor changes in risk over time (Douglas & Skeem, 2005; Mills, 2017).

The present study added to the growing body of empirical research that can speak to the internal and external responsiveness of dynamic risk ratings on the HCR-20

and START. Overall, the available empirical research supports the internal responsiveness of the dynamic ratings on these tools. That is, the dynamic risk and protective factors, dynamic scale scores, and summary risk judgments are capable of intraindividual change over time. The available empirical research supporting the external responsiveness of these ratings is more limited, yet studies have found support for a relationship between change on the dynamic ratings on these tools and future violence perpetration. As such, the items on the HCR-20 and START are appropriately labelled as dynamic risk and protective factors according to contemporary definitions (Andrews & Bonta, 2010b; Douglas & Skeem, 2005).

The continued focus on dynamic risk factors is vital, as comprehensive violence risk assessment tools must facilitate the identification of treatment targets and monitoring of risk over time. Research must continue to examine dynamic risk ratings. Further research is needed to better understand which risk factors are best characterized as dynamic and/or causal risk factors. These risk factors have the greatest potential for violence prevention. Future research should investigate the defining features of different types of risk factors, including their ability to change over time and their association with violence. Research should continue to examine various characteristics of change in dynamic risk ratings. Future research should examine the temporal precedence of sets of dynamic risk factors, the interrelationships between dynamic risk factors over time, and the individual and combined association between dynamic risk factors and violence over time. It is still the case that research is needed “to disentangle the independent, indirect, interactive, or transactional effects of dynamic risk factors on violence” (Douglas & Skeem, 2005, p. 368).

References

Adedokun, O. A., & Burgess, W. D. (2012). Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of Multidisciplinary Evaluation*, *8*(17), 125–131.

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Ahn, C. (2008). Sample size requirements for clinical trials with repeated binary outcomes. *Drug Information Journal*, *42*, 107–113.

Ainsworth, S. A., & Taxman, F. S. (2013). Creating simulation parameter inputs with existing data sources: Estimating offender risks, needs, and recidivism. In F. S. Taxman & A. Pattavina (Eds.), *Simulation strategies to reduce recidivism: Risk Need Responsivity (RNR) modeling in the criminal justice system* (pp. 115–142). Springer.

Andrews, D. A. (2012). The risk-need-responsivity (RNR) model of correctional assessment and treatment. In J. A. Dvoskin, J. L. Skeem, R. W. Novaco, & K. S. Douglas (Eds.), *Using social science to reduce violent offending* (pp. 127–156). Oxford University Press.

Andrews, D. A., & Bonta, J. (2010a). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy and Law*, *16*, 39–55. https://doi.org/10.1037/a0018362

Andrews, D. A., & Bonta, J. (2010b). *The psychology of criminal conduct* (5th ed.). Anderson.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, *17*, 19–22. https://doi.org/10.1177/0093854890017001004

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. https://doi.org/10.3138/cjccj.49.4.439

Arai, K., Takano, A., Nagata, T., & Hirabayashi, N. (2017). Predictive accuracy of the Historical-Clinical-Risk Management-20 for violence in forensic psychiatric wards in Japan: The HCR-20 in Japanese forensic psychiatric wards. *Criminal Behaviour and Mental Health*, *27*(5), 409–420. https://doi.org/10.1002/cbm.2007

Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system*. National Council on Crime and Delinquency.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456

Barefoot v. Estelle, 463 U.S. 880 (1983).

Barnett, G. D., Wakeling, H. C., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? Psychology, *Crime & Law*, *19*(1), 85–110. http://dx.doi.org/10.1080/1068316X.2011.607820

Bartko, J. J., & Carpenter, W. T. (1976). On the methods and theory of reliability. *Journal of Nervous and Mental Disease*, *163*(5), 307–317. https://doi.org/10.1097/00005053-197611000-00003

Beaton, D. E., Bombardier, C., Katz, J. N., & Wright, J. G. (2001). A taxonomy for responsiveness. *Journal of Clinical Epidemiology*, *54*(12), 1204–1217. https://doi.org/10.1016/s0895-4356(01)00407-3

Beaton, D. E., Hogg-Johnson, S., & Bombardier, C. (1997). Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *Journal of Clinical Epidemiology*, *50*(1), 79–93. https://doi.org/10.1016/S0895-4356(96)00296-X

Beckerman H., Roebroeck M. E., Lankhorst G. J., Becher J. G., Bezemer P. D., & Verbeek A. L. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research*, *10*, 571–578. https://doi.org/10.1023/a:1013138911638

Beckerman H., Vogelaar T. W., Lankhorst G. J., & Verbeek A. L. (1996). A criterion for stability of the motor function of the lower extremity in stroke patients using the Fugl–Meyer Assessment Scale. *Scandinavian Journal of Rehabilitation Medicine*, *28*(1), 3–7.

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. https://doi.org/10.1177/107906320201400206

Beech, A. R., Wakeling, H. C., Szumski, F., & Freemantle, N. (2016). Problems in the measurement of dynamic risk factors in sexual offenders. *Psychology, Crime & Law*, *22*(1-2), 68–83. https://doi.org/10.1080/1068316X.2015.1109095

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. https://doi.org/10.1037/a0022900

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. https://doi.org/10.1080/14999013.2002.10471158

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. https://doi.org/10.1080/095851800362445

Belfrage, H., Fransson, G., & Strand, S. (2004). Management of violent behaviour in the correctional system using qualified risk assessments. *Legal and Criminological Psychology*, *9*(1), 11–22. https://doi.org/10.1348/135532504322776825

Bjørkly, S., Eidhammer, G., & Selmer, L. E. (2014). Concurrent validity and clinical utility of the HCR-20^{V3} compared with the HCR-20 in forensic mental health nursing: Similar tools but improved method. *Journal of Forensic Nursing*, *10*(4), 234–242. https://doi.org/10.1097/JFN.0000000000000047

Blanchard, A. J. E. (2013). *Dynamic risk factors in violence risk assessment: A multiple time-point evaluation of the HCR-20 and START* [Unpublished master's thesis]. Simon Fraser University.

Blanchard, A. J. E., Reeves, K. A., & Gibas, A. L. (2016). Canadian contributions to violence risk assessment: Policy, practice, and future directions. In J. P. Singh, S Bjørkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment* (pp. 202–215). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199386291.003.0013

Bland, J. M., & Altman, D. G. (1995a). Statistical notes: Calculating correlation coefficients with repeated observations: Part 1 – Correlation within subjects. *BMJ*, *310*, 446. https://doi.org/10.1136/bmj.310.6977.446

Bland, J. M., & Altman, D. G. (1995b). Calculating correlation coefficients with repeated observations: Part 2 – Correlation between subjects. *BMJ*, *310*, 633. https://doi.org/10.1136/bmj.310.6980.633

Bland, J. M., & Altman, D. G. (1996). Statistics notes: Measurement error. *BMJ*, *313*(7059), 744–744. https://doi.org/10.1136/bmj.313.7059.744

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

Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior*, *29*(4), 355–379. https://doi.org/10.1177/0093854802029004002

Bonta, J. L., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). Routledge.

Bowen, E., Gilchrist, E., & Beech, A. R. (2008). Change in treatment has no relationship with subsequent re-offending in u.k. domestic violence sample: A preliminary study. *International Journal of Offender Therapy and Comparative Criminology*, *52*(5), 598–614. https://doi.org/10.1177/0306624X08319419

Brame, B., Nagin, D. S., & Tremblay, R. E. (2001). Developmental trajectories of physical aggression from school entry to late adolescence. *Journal of Child Psychiatry*, *42*(4), 503–512. https://doi.org/10.1111/1469-7610.00744

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). *The International Journal of Forensic Mental Health*, *9*(4), 271–281. https://doi.org/10.1080/14999013.2010.534378

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*(1), 25–45. https://doi.org/10.1007/s10979-008-9139-7

Bujang, M. A., & Baharum, N. (2016). Sample size guideline for correlation analysis. *World Journal of Social Science Research*, *3*(1), 37–46. https://doi.org/10.22158/wjssr.v3n1p37

Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications*. Chapman and Hall. https://doi.org/10.1201/9781420066586

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. https://doi.org/10.1177/0093854809333610

Canadian Charter of Rights and Freedoms, s 2, Part I of the *Constitution Act*, 1982, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11.

Casey, S. (2016). Dynamic risk and sexual offending: The conundrum of assessment. *Psychology, Crime & Law*, *22*(1-2), 104–123. https://doi.org/10.1080/1068316X.2015.1111366

Caudy, M. S., Durso, J. M., & Taxman, F. S. (2013). How well do dynamic needs predict recidivism? Implications for risk assessment and risk reduction. *Journal of Criminal Justice*, *41*(6), 458–466. https://doi.org/10.1016/j.jcrimjus.2013.08.004

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. https://doi.org/10.1080/14999013.2011.629715

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*(2), 127–137.

Coid, J. W., Kallis, C., Doyle, M., Shaw, J., & Ullrich, S. (2015). Identifying causal risk factors for violence among discharged patients. *PLOS ONE*, *10*(11), e0142493. https://doi.org/10.1371/journal.pone.0142493

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). Routledge.

Cording, J. R., Beggs Christofferson, S. M., & Grace, R. C. (2016). Challenges for the theory and application of dynamic risk factors. *Psychology, Crime & Law*, *22*(1-2), 84–103. https://doi.org/10.1080/1068316X.2015.1111367

Coupland, R. B. A., & Olver, M. E. (2018). Assessing dynamic violence risk in a high-risk treated sample of violent offenders. *Assessment*, *27*(8), 1886–1900. https://doi.org/10.1177/1073191118797440

Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression*, *11*(1), 65–84. https://doi.org/10.1080/13552600410001667733

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. https://doi.org/10.1080/355260031000137968

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. *The International Journal of Forensic Mental health*, *6*(2), 137–143. https://doi.org/10.1080/14999013.2007.10471258

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. https://doi.org/10.1080/14789940903173990

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. https://doi.org/10.1177/107906320201400204

Desmarais, S. L., Nicholls, T., 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. https://doi.org/10.1037/a0026668

de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime, & Law*, *12*(3), 321–336. https://doi.org/10.1080/10683160600569029

de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *The International Journal of Forensic Mental Health*, *3*(2), 149–165. https://doi.org/10.1080/14999013.2004.10471204

de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. I. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in postdischarge community recidivism. *Law and Human Behavior*, *39*(1), 53–61. https://doi.org/10.1037/lhb0000089

Dexter, E., & Vitacco, M. J. (2020). Strategies for assessing and preventing inpatient violence in forensic hospitals: A call for specificity. *European Psychologist*, *25*(2), 146–154. https://doi.org/10.1027/1016-9040/a000393

Diggle P. J., Heagerty, P., Liang K.-Y., & Zeger S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford University 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. https://doi.org/10.1080/14789940701489390

Douglas, K. S., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 version 3. *International Journal of Forensic Mental Health*, *13*(2), 130–139. https://doi.org/10.1080/14999013.2014.908429

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). Routledge.

Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology*, *4*, 149–184. https://doi.org/10.1348/135532599167824

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20 (Version 3): Assessing risk of violence-User guide*. Mental Health Law and 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. https://doi.org/10.1177/009385402236735

Douglas, K. S., & Ogloff, J. R. P. (2003). Multiple facets of risk for violence: The impact of judgmental specificity on structured decisions about violence risk. *The International Journal of Forensic Mental Health*, *2*(1), 19–34. https://doi.org/10.1080/14999013.2003.10471176

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. https://doi.org/10.1176/appi.ps.54.10.1372

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. https://doi.org/10.1037//0022-006x.67.6.917

Douglas, K. S., & Otto, R. K. (Eds.). (2020). *Handbook of violence risk assessment* (2nd ed.). Routledge. https://doi.org/10.4324/9781315518374-19

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). Routledge.

Douglas, K. S., & Shaffer, C. S. (2020). The Science of and practice with the HCR-20 V3 (Historical-Clinical-Risk Management-20, Version 3). In K. S. Douglas & R. K. Otto (Eds.), *Handbook of Violence Risk Assessment* (2nd ed., pp. 253–293). Routledge. https://doi.org/10.4324/9781315518374-16

Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K., & Weir, J. (2014). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. HCR-20 Violence Risk Assessment White Paper Series, #1. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*, *11*(3), 347–383. https://doi.org/10.1037/1076-8971.11.3.347

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. https://doi.org/10.1177/0093854805278411

Dowden, C., & Andrews, D. A. (1999a). What works for female offenders: A meta-analytic review. *Crime and Delinquency*, *45*, 438–452. https://doi.org/10.1177/0011128799045004002

Dowden, C., & Andrews, D. A. (1999b). What works in young offender treatment: A meta-analysis. *Forum on Corrections Research*, *11*(2), 21–24. https://www.csc-scc.gc.ca/research/forum/e112/112e_e.pdf

Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, *42*(4), 449–467. https://doi.org/10.3138/cjcrim.42.4.449

Dowden, C., & Andrews, D. A. (2003). Does family intervention work for delinquents? Results of a meta-analysis. Can*adian Journal of Criminology and Criminal Justice*, *45*(3), 327–342. https://doi.org/10.3138/cjccj.45.3.327

Dowden, C., & Andrews, D. A. (2004). The importance of staff practice in delivering effective correctional treatment: A meta-analytic review of core correctional practice. *International Journal of Offender Therapy and Comparative Criminology*, *48*(2), 203–214. https://doi.org/10.1177/0306624X03257765

Doyle, M., Lewis, G., & Brisbane, M. (2008). Implementing the Short-Term Assessment of Risk and Treatability (START) in a forensic mental health service. *Psychiatric Bulletin*, *32*(11), 406–408. https://doi.org/10.1192/pb.bp.108.019794

Doyle, M., & Logan, C. (2012). Operationalizing the assessment and management of violence risk in the short-term: Short-term risk management. *Behavioral Sciences & the Law*, *30*(4), 406–419. https://doi.org/10.1002/bsl.2017

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. https://doi.org/10.1080/14789949.2010.481720

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology*, *27*(3), 248–261. https://doi.org/10.1093/arclin/acr120

Duffy, S. W. (1984). Asymptotic and exact power for the McNemar test and its analogue with R controls per case. *Biometrics*, *40*(4), 1005–1015. https://doi.org/10.2307/2531151

Durrant, R. (2016). Putting risk factors in their place: An evolutionary-developmental approach to understanding risk. Psychology, Crime & Law, 22(1-2), 17–32. https://doi.org/10.1080/1068316X.2015.1109093

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., Skeem, J. S., Novaco, R., & Douglas, K. S. (Eds.). (2012). *Using social science to reduce crime*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195384642.001.0001

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. https://doi.org/10.1176/ps.45.7.679

Doyle, M., & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services. *British Journal of Psychiatry*, *189*(6), 520–526. https://doi.org/10.1192/bjp.bp.105.021204

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). 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. https://doi.org/10.1177/1079063211403164

Eisenberg, M. J., van Horn, J. E., Dekker, J. M., Assink, M., van der Put, C. E., Hendriks, J., & Stams, G. J. J. M. (2019). Static and dynamic predictors of general and violent criminal offense recidivism in the forensic outpatient population: A meta-analysis. *Criminal Justice and Behavior*, *46*(5), 732–750. https://doi.org/10.1177/0093854819826109

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabilitation Psychology*, *56*(4), 267–288. https://doi.org/10.1037/a0025579

Ennis, B. J., & Litwack, T. R. (1974). Psychiatry and the presumption of expertise: Flipping coins in the courtroom. *California Law Review*, *62*(3), 693–752.

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidenced Based Mental Health, 1,* 70–72. https://doi.org/10.1136/ebmh.1.3.70

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24,827 people: Systematic review and meta-analysis. *BMJ*, *345*, e4692. https://doi.org/10.1136/bmj.e4692

Freeman, J., Walters, R., Ingram, W., Slade, A., Hobart, J., & Zajicek, J. (2013). Evaluating change in mobility in people with multiple sclerosis: Relative responsiveness of four clinical measures. *Multiple Sclerosis Journal*, *19*(12), 1632–1639. https://doi.org/10.1177/1352458513482373

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

Garson, G. D. (2015). *Missing values and data imputation* (2015 edition). Blue Book Series. Statistical Associates Publishing.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, *34*(4), 575–608. https://doi.org/10.1111/j.1745-9125.1996.tb01220.x

Gibbons, R. D., Hedeker, D., & DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology*, *6*(1), 79–107. https://doi.org/10.1146/annurev.clinpsy.032408.153550

Grann, M., Sturidsson, K., Haggård-Grann, U., Hiscoke, U. L., Alm, P.-O., Dernevik, M., Gumpert, C., Hallqvist, J., Hallquist, T., Kullgren, G., Långström, N., Lotterberg, M., Nordström, K., Ståhle, B., & Woodhouse, A. (2005). Methodological development: Structured outcome assessment and community risk monitoring (SORM). *International Journal of Law and Psychiatry*, *28*(4), 442–456. https://doi.org/10.1016/j.ijlp.2004.03.011

Gray, N. S., Benson, R., Craig, R., Davies, H., Fitzgerald, S., Huckle, P., Maggs, R., Taylor, J., Trueman, M., Williams, T., & Snowden, R. J. (2011). The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *The International Journal of Forensic Mental Health*, *10*(4), 305–313. https://doi.org/10.1080/14999013.2011.631692

Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology*, *71*(3), 443–451. https://doi.org/10.1037/0022-006X.71.3.443

Greiner, L. E., Law, M. A., & Brown, S. L. (2015). Using dynamic factors to predict recidivism among women: A four-wave prospective study. *Criminal Justice and Behavior*, *42*(5), 457–480. https://doi.org/10.1177/0093854814553222

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. https://doi.org/10.1002/mpr.330

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

Guy, L. S., Douglas, K. S., & Hart, S. D. (2015). Risk assessment and communication. In B. L. Cutler, & P. A. Zapf (Eds.), *APA handbook of forensic psychology: Volume 1. Individual and situational influences in criminal and civil contexts* (pp. 35–86). American Psychological Association.

Guyatt, G. H., Walter, S. D., & Norman, G. R. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. Journal of Chronic Diseases, 40(2), 171–178. https://doi.org/10.1016/0021-9681(87)90069-5

Guyatt, G. H. Van Zanten, S. J. O. V., Feeny, D. H., & Patrick, D. L. (1989). Measuring quality of life in clinical trials; A taxonomy and review. *CMAJ*, *140*(12), 1441–1448.

Haggård-Grann, U. (2007). Assessing violence risk: A review and clinical recommendations. *Journal of Counselling & Development*, *85*, 294–301. https://doi.org/10.1002/j.1556-6678.2007.tb00477.x

Hannah-Moffat, K. (2004). Losing ground: Gendered knowledges, parole risk, and responsibility. *Social Politics: International Studies in Gender, State and Society*, *11*, 363–385.

Hannah-Moffat, K. (2009). Gridlock or mutability: Reconsidering "gender" and risk assessment. *Criminology and Public Policy*, *8*(1), 209–219. https://doi.org/10.1111/j.1745-9133.2009.00549.x

Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law*, *4*(1-2), 50–72. https://doi.org/10.1037/1076-8971.4.1-2.50

Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, *20*(2), 212–217. https://doi.org/10.1177/0886260504267740

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior*, *27*(1), 6–35. https://doi.org/10.1177/0093854800027001002

Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, (2007). *Assessing the risk of sexual offenders on community supervision: The dynamic supervision project*. Public Safety Canada. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/index-en.aspx

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, *21*(1), 1–21. https://doi.org/10.1037/a0014421

Hare, R. (1991). *The Hare Psychopathy Checklist-Revised*. Multi-Health Systems.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). *Violent offenders: Appraising and managing risk* (3rd ed.). American Psychological Association. https://doi.org/10.1037/14572-000

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, *3*, 121–137. https://doi.org/10.1111/j.2044-8333.1998.tb00354.x

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

Hart, S. D. (2008). Preventing violence: The role of 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). Nova Science.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare Psychopathy Checklist: Screening Version* (1st ed.). Multi-Health Systems.

Hart, S. D., Douglas, K. S., & Webster, C. D. (2001). Risk management using the HCR-20: A general overview focusing on historical factors. 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. 27–40). Mental Health, Law, and Policy Institute, Simon Fraser University.

Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgment approach. In P. Sturmey & M. McMurran (Eds.), *Forensic case formulation* (pp. 83–106). Wiley-Blackwell.

Haynes, S. N. (1992). *Models of causality in psychopathology: Toward synthetic, dynamic, and nonlinear models of causality in psychopathology*. Allyn & Bacon.

Heffernan, R., & Ward, T. (2015). The conceptualization of dynamic risk factors in child sex offenders: An agency model. *Aggression and Violent Behavior*, *24*, 250–260. https://doi.org/10.1016/j.avb.2015.07.001

Heffernan, R., & Ward, T. (2017). A comprehensive theory of dynamic risk and protective factors. *Aggression and Violent Behavior*, *37*, 129–141. https://doi.org/10.1016/j.avb.2017.10.003

Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, *21*(4), 347–359. https://doi.org/10.1023/A:1024851017947

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). Guilford Press.

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*(1-2), 138–174. https://doi.org/10.1037/1076-8971.4.1-2.138

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*(4), 276–288. https://doi.org/10.1016/j.ijlp.2012.04.001

Hinton-Bayre, A. D. (2010). Deriving reliable change statistics from test-retest normative data: Comparison of models and mathematical expressions. *Archives of Clinical Neuropsychology*, *25*(3), 244–256. https://doi.org/10.1039/arclin/acq008

Hogan, N., & Ennis, L. (2010). Assessing risk for forensic psychiatric inpatient violence: A meta-analysis. *Open Access Journal of Forensic Psychology*, *2*, 137–147.

Hogan, N. R., & Olver, M. E. (2016). Assessing risk for aggression in forensic psychiatric inpatients: An examination of five measures. *Law and Human Behavior*, *40*(3), 233–243. https://doi.org/10.1037/lhb0000179

Hogan, N. R., & Olver, M. E. (2018). A prospective examination of the predictive validity of five structured instruments for inpatient violence in a secure forensic hospital. *International Journal of Forensic Mental Health*, *17*(2), 122–132. https://doi.org/10.1080/14999013.2018.1431339

Hogan, N. R., & Olver, M. E. (2019). Static and dynamic assessment of violence risk among discharged forensic patients. *Criminal Justice and Behavior*, *46*(7), 923–938. https://doi.org/10.1177/0093854819846526

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 & the Law*, *30*(4), 431–447. https://doi.org/10.1002/bsl.2016

Howard, P. D., & Dixon, L. (2013). Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys violence predictor. *Law and Human Behavior*, *37*(3), 163–174. https://doi.org/10.1037/lhb0000012

Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment*, *15*(4), 443–445. https://doi.org/10.1037/1040-3590.15.4.443

Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment*, *11*(3), 266–277. https://doi.org/10.1037/1040-3590.11.3.266

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology*, *53*, 459–468. https://doi.org/10.1016/s0895-4356(99)00206-1

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. https://doi.org/10.1037//0022-006x.59.1.12

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. https://doi.org/10.1177/0093854810368924

Jung, S.-H., & Ahn, C. (2003). Sample size estimation for GEE method for comparing slopes in repeated measurements data. *Statistics in Medicine*, *22*(8), 1305–1315. https://doi.org/10.1002/sim.1384

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. https://doi.org/10.1016/s0272-7358(97)00012-3

Kingston, D. A., & Olver, M. E. (2018). Psychometric examination of treatment change among mentally disordered offenders: A risk–needs analysis. *Criminal Justice and Behavior*, *45*(2), 153–172. https://doi.org/10.1177/0093854817743539

Klepfisz, G., Daffern, M., & Day, A. (2016). Understanding dynamic risk factors for violence. *Psychology, Crime & Law*, *22*(1-2), 124–137. https://doi.org/10.1080/1068316X.2015.1109091

Kochenderfer-Ladd, B., & Wardrop, J. L. (2001). Chronicity and instability of children's peer victimization experiences as predictors of loneliness and social satisfaction trajectories. *Child Development*, *72*(1), 134–151. https://doi.org/10.1111/1467-8624.00270

Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology*, *9*(1), 19–43. https://doi.org/10.1007/s11292-012-9159-7

Kozol, H. L., Boucher, R. J., & Garofalo, R. F. (1972). The diagnosis and treatment of dangerousness. *Crime and Delinquency*, *18*, 371–392. https://doi.org/10.1177/001112877201800407

Kraemer, H. C. (2003). Current concepts of risk in psychiatric disorders. *Current Opinion in Psychiatry*, *16*(4), 421–430. https://doi.org/10.1097/01.yco.0000079210.36371.52

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*(4), 337–343. https://doi.org/10.1001/archpsyc.1997.01830160065009

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. https://doi.org/10.1176/appi.ajp.158.6.848

Kroner, D. G., & Yessine, A. K. (2013). Changing risk factors that impact recidivism: In search of mechanisms of change. *Law and Human Behavior*, *37*(5), 321–336. https://doi.org/10.1037/lhb0000022

Lachin, J. M. (1992). Power and sample size evaluation for the McNemar test with application to matched case-control studies. *Statistics in Medicine*, *11*(9), 1239–1251. https://doi.org/10.1002/sim.4780110909

Lambert, M. J., & Vermeersch, D. A. (2013). Psychological assessment in treatment. In M. C. Rodriguez (Ed.), *APA handbook of testing and assessment in psychology: Vol. 2. Testing and assessment in clinical and counseling psychology* (pp. 213–229). American Psychological Association. https://doi.org/10.1037/14048-013

Landis, J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lavoie, J. A. A., Guy, L. A., & 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. 3–19). Willan.

Leon, A. C. (2004). Sample-size requirements for comparisons of two groups on repeated observations of a binary outcome. *Evaluation & the Health Professions*, *27*(1), 34–44. https://doi.org/10.1177/0163278703261198

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. https://doi.org/10.1177/1073191112441242

Li, Z., & McKeague, I. W. (2013). Power and sample size calculations for generalized estimating equations via local asymptotics. *Statistica Sinica*, *23*(1), 231–250. https://doi.org/10.5705/ss.2011.081

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Litwack, T. R., & Schlesinger, L. B. (1999). Dangerousness risk assessments: Research, legal, and clinical considerations. In A. K. Hess & I. B. Weiner (Eds.), *The handbook of forensic psychology* (2nd ed., pp. 171–217). Wiley.

Liu, G., & Liang, K.-Y. (1997). Sample size calculations for studies with correlated observations. *Biometrics*, *53*(3), 937–947. https://doi.org/10.2307/2533554

Lloyd, C. D., Hanby, L. J., & Serin, R. C. (2014). Rehabilitation group coparticipants' risk levels are associated with offenders' treatment performance, treatment change, and recidivism. *Journal of Consulting and Clinical Psychology*, *82*(2), 298–311. https://doi.org/10.1037/a0035360

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. https://doi.org/10.1037/a0024388

Lyon, D. R., Hart, S. D., & Webster, C. D. (2001). Violence and risk assessment. In R. A. Schuller & J. R. P. Ogloff (Eds.), *Introduction to psychology and law: Canadian perspectives* (pp. 314–350). University of Toronto Press.

Lyons v. the Queen, 37 C.C.C. (3d) 1 at 47 (S.C.C.) (1987).

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, *22*(2), 191–217. https://doi.org/10.1177/1079063210366039

Martin, T., & Daffern, M. (2006). Clinician perceptions of personal safety and confidence to manage inpatient aggression in a forensic psychiatric setting. *Journal of Psychiatric and Mental Health Nursing*, *13*, 90–99. https://doi.org/10.1111/j.1365-2850.2006.00920.x

Mastromanno, B., Brookstein, D. M., Ogloff, J. R. P., Campbell, R., Chu, C. M., & Daffern, M. (2018). Assessing change in dynamic risk factors in forensic psychiatric inpatients: Relationship with psychopathy and recidivism. *The Journal of Forensic Psychiatry & Psychology*, *29*(2), 323–336. https://doi.org/10.1080/14789949.2017.1377277

May, J. O., & Looney, S. W. (2020). Sample size charts for Spearman and Kendall coefficients. *Journal of Biometrics & Biostatistics*, *11*(2). https://doi.org/10.37421/jbmbs.2020.11.440

McDermott, B. E., Quanbeck, C. D., Busse, D., Yastro, K., & Scott, C. L. (2008). The accuracy of risk assessment instruments in the prediction of impulsive versus predatory aggression. *Behavioral Sciences & the Law*, *26*(6), 759–777. https://doi.org/10.1002/bsl.842

McDermott, S., & Nagin, D. S. (2001). Same or different? Comparing offender groups and covariates over time. *Sociological Methods & Research*, *29*(3), 282–318. https://doi.org/10.1177/0049124101029003002

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*(4), 529–550. https://doi.org/10.1016/S0005-7894(02)80015-6

McGlinchey, J. B., & Jacobson, N. S. (1999). Clinically significant but impractical? A response to Hageman and Arrindell. *Behaviour Research and Therapy*, *37*, 1211-1217. https://doi.org/10.1016/S0005-7967(99)00035-2

McNaughton Nicholls, C. & Webster, S. (2014). *Sex offender management and dynamic risk: Pilot evaluation of the Active Risk Management System (ARMS)*. Ministry of Justice Analytical Series, National Offender Management Services.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157. https://doi.org/10.1007/BF02295996

McNiel, D. E. (1998). Empirically based clinical evaluation and management of the potentially violent patient. In P. M. Kleespies (Ed.), *Emergencies in mental health practice: Evaluation and management*. (pp. 95–116). Guilford Press.

McNiel, D. E., Borum, R., Douglas, K. S., Hart, S. D., Lyon, D., Sullivan, L. E., & Hemphill, J. F. (2002). Risk Assessment. In J. R. P. Ogloff (Ed.), *Taking psychology and law into the twenty-first century* (pp. 147–170). Kluwer Academic/Plenum Publishers.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (3rd ed.). Guilford Press.

Michel, S. F., Riaz, M., Webster, C., Hart, S. D., Levander, S., Müller-Isberner, R., Tiihonen, J., Repo-Tiihonen, E., Tuninger, E., & Hodgins, S. (2013). Using the HCR-20 to predict aggressive behavior 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. https://doi.org/10.1080/14999013.2012.760182

Mills, J. F. (2005). Advances in the assessment and prediction of interpersonal violence. *Journal of Interpersonal Violence*, *20*(2), 236–241. https://doi.org/10.1177/0886260504267745

Mills, J. F. (2017). Violence risk assessment: A brief review, current issues, and future directions. *Canadian Psychology/Psychologie Canadienne*, *58*(1), 40–49. https://doi.org/10.1037/cap0000100

Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Sage Publications.

Monahan, J. (1984). The Prediction of violent behavior: Toward a second generation of theory and policy. *American Journal of Psychiatry*, *141*(1), 10–15. https://doi.org/10.1176/ajp.141.1.10

Monahan, J. (1993). Limiting therapist exposure to *Tarasoff* liability. *American Psychologist*, *48*(3), 242–250. https://doi.org/10.1037/0003-066X.48.3.242

Monahan, J. (1996). Violence prediction: The last 20 years and the next 20 years. *Criminal Justice and Behavior*, *23*(1), 107–120.

Monahan, J., & Skeem, J. L. (2014). Risk Redux. *Federal Sentencing Reporter*, *26*(3), 158–166. https://doi.org/10.1525/fsr.2014.26.3.158

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, *12*(1), 489–513. https://doi.org/10.1146/annurev-clinpsy-021815-092945

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). American Psychological Association. https://doi.org/10.1037/10414-010

Monahan, J., Steadman, H. J., Silver, E., Applebaum, 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*. Oxford University Press.

Morrissey, C., Beeley, C., & Milton, J. (2014). Longitudinal HCR-20 scores in a high-secure psychiatric hospital. *Criminal Behaviour and Mental Health*, *24*(3), 169–180. https://doi.org/10.1002/cbm.1893

Müller-Isberner, R., Webster, C. D., & Gretenkord, L. (2007). Measuring progress in hospital order treatment: Relationship between levels of security and C and R Scores of the HCR-20. *International Journal of Forensic Mental Health*, *6*(2), 113–121. https://doi.org/10.1080/14999013.2007.10471256

Murrie, D. C., & Kelley, S. (2017). Evaluating and managing the risk of violence in clinical practice with adult. In P. M. Kleespies (Ed.), *The Oxford handbook of behavioral emergencies and crises* (pp. 126–145). Oxford University Press.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*(2), 139–157. https://doi.org/10.1037/1082-989X.4.2.139

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. https://doi.org/10.1111/1467-8624.00086

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. https://doi.org/10.1177/1094428103254673

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. https://doi.org/10.1177/1073191106290559

Nicholls, T. L., Cruise, K. R., Greig, D., & Hinz, H. (2015). Female offenders. In B. L. Cutler, & P. A. Zapf (Eds.), *APA handbook of forensic psychology: Volume 2. Criminal investigation, adjudication, and sentencing outcomes* (pp. 79–123). American Psychological Association.

Nicholls, T. L., Petersen, K., Almond, M., & Geddes, C. (2021). *Short-Term Assessment of Risk and Treatability (START): Rationale, Application, and empirical overview.* In K. S. Douglas & R. K. Otto (Eds.), *Handbook of Violence Risk Assessment* (2nd ed., pp. 359–388). Routledge. https://doi.org/10.4324/9781315518374-19

Nitschke, J., Sünkel, Z., & Mokros, A. (2020a). Die forensische Präventionsambulanz Ansbach: Evaluation des Modellprojekts zur Behandlung psychiatrischer Risikopatienten. *Der Nervenarzt*, *91*(5), 439–445. https://doi.org/10.1007/s00115-019-0752-0

Nitschke, J., Sünkel, Z., & Mokros, A. (2020b). Forensic prevention outpatient clinic: A treatment evaluation study. *The Journal of Forensic Psychiatry & Psychology*, *31*(5), 714–730. https://doi.org/10.1080/14789949.2020.1785527

Nonstad, K., Nesset, M. B., Kroppan, E., Pedersen, T. W., Nøttestad, J. A., Almvik, R., & Palmsteirna, T. (2010). Predictive validity and other psychometric properties of the Short-Term Assessment of Risk and Treatability (START) in Norwegian high secure hospital. *International Journal of Forensic Mental Health*, *9*, 74–81. https://doi.org/10.1080/14999013.2010.534958

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. https://doi.org/10.1176/appi.ajp.2008.08091398

Olsson, H., Strand, S., Kristiansen, L., Sjöling, M., & Asplund, K. (2013). Decreased risk for violence in patients admitted to forensic care, measured with the HCR-20. *Archives of Psychiatric Nursing*, *27*(4), 191–197. https://doi.org/10.1016/j.apnu.2013.03.004

Olver, M. E., Nicholaichuk, T. P., & Wong, S. C. P. (2014). The predictive and convergent validity of a psychometric battery used to assess sexual offenders in a treatment programme: An 18-year follow-up. *Journal of Sexual Aggression*, *20*(2), 216–239. https://doi.org/10.1080/13552600.2013.816791

Olver, M. E., & Wong, S. C. P. (2011). A comparison of static and dynamic assessment of sexual offender risk and need in a treatment context. *Criminal Justice and Behavior*, *38*(2), 113–126. https://doi.org/10.1177/0093854810389534

Olver, M. E., & Stockdale, K. C. (2020). Evaluating change in men who have sexually offended: Linkages to risk assessment and management. *Current Psychiatry Reports*, *22*(5), 22. https://doi.org/10.1007/s11920-020-01146-3

Osgood, D. W., & Smith, G. L. (1995). Applying hierarchical linear modeling to extended longitudinal evaluations: The boys town follow-up study. *Evaluation Review*, *19*(1), 3–38. https://doi.org/10.1177/0193841X9501900101

O'Shea, L. E., & Dickens, G. L. (2014). Short-Term Assessment of Risk and Treatability (START): Systematic review and meta-analysis. *Psychological Assessment*, *26*(3), 990–1002. https://doi.org/10.1037/a0036794

O'Shea, L. E., & Dickens, G. L. (2015). The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients. *Comprehensive Psychiatry*, *62*, 132–140. https://doi.org/10.1016/j.comppsych.2015.07.009

O'Shea, L. E., Mitchell, A. E., Picchioni, M. M., & Dickens, G. L. (2013). Moderators of the predictive efficacy of the Historical, Clinical and Risk Management-20 for aggression in psychiatric facilities: Systematic review and meta-analysis. *Aggression and Violent Behavior*, *18*(2), 255–270. https://doi.org/10.1016/j.avb.2012.11.016

Otto, R. K. (2000). Assessing and managing violence risk in outpatient settings. *Journal of Clinical Psychology*, *56*(10), 1239–1262. https://doi.org/10.1002/1097-4679(200010)56:10<1239::AID-JCLP2>3.0.CO;2-J

Pedersen, L., Rasmussen, K., & Elsass, P. (2010). Risk Assessment: The Value of Structured Professional Judgments. *International Journal of Forensic Mental Health*, *9*(2), 74–81. https://doi.org/10.1080/14999013.2010.499556

Penney, S. R., Marshall, L. A., & Simpson, A. I. F. (2016). The assessment of dynamic risk among forensic psychiatric patients transitioning to the community. *Law and Human Behavior*, *40*(4), 374–386. https://doi.org/10.1037/lhb0000183

Polaschek, D. L. L., Yesberg, J. A., Bell, R. K., Casey, A. R., & Dickson, S. R. (2016). Intensive psychological treatment of high-risk violent offenders: Outcomes and pre-release mechanisms. *Psychology, Crime & Law*, *22*(4), 344–365. https://doi.org/10.1080/1068316X.2015.1109088

Privitera, G. J. (2018). *Statistics for the behavioral sciences* (3rd ed.). Sage Publications.

Quinsey, V., Harris, G., Rice, M., & Cormier, C. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). 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. https://doi.org/10.1177/0886260506294238

Raghunathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods*, *1*(1), 178-183. https://doi.org/10.1037/1082-989X.1.2.178

Ramesh, T., Igoumenou, A., Vazquez Montes, M., & Fazel, S. (2018). Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: A systematic review and meta-analysis. *European Psychiatry*, *52*, 47–53. https://doi.org/10.1016/j.eurpsy.2018.02.007

Riddle, D., & Stratford, P. (2013). *Is this change real? Interpreting patient outcomes in physical therapy*. F. A. Davis Company.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*(2), 117–141. https://doi.org/10.1037/bul0000088

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Rossdale, S.-V., Tully, R. J., & Egan, V. (2020). The HCR-20 for predicting violence in adult females: A meta-analysis. *Journal of Forensic Psychology Research and Practice*, *20*(1), 15–52. https://doi.org/10.1080/24732850.2019.1681875

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons. https://doi.org/10.1002/9780470316696

Salekin, R.T. (2002). Psychopathy and therapeutic pessimism: Clinical lore or clinical reality? *Clinical Psychology Review*, *22*, 79–112. https://doi.org/10.1016/s0272-7358(01)00083-6

Schlesinger, L. B. (1996). The Catathymic crisis, 1912-present: A review and clinical study. *Aggression and Violent Behavior*, *1*(4), 307–316. https://doi.org/10.1016/S1359-1789(96)00003-1

Schuck, P., & Zwingmann, C. (2003). The 'smallest real difference' as a measure of sensitivity to change: A critical analysis: *International Journal of Rehabilitation Research*, *26*(2), 85–91. https://doi.org/10.1097/01.mrr.0000070759.63544.65

Scott, P. D. (1977). Assessing dangerousness in criminals. *The British Journal of Psychiatry*, *131*, 127–142. https://doi.org/10.1192/bjp.131.2.127

Scurich, N. (2016). An introduction to the assessment of violence risk. In J. P. Singh, S Bjørkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment* (pp. 3–15). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199386291.003.0001

Sellers, B. G., Desmarais, S. L., & Hanger, M. W. (2017). Measurement of change in dynamic factors using the START: AV. *Journal of Forensic Psychology Research and Practice*, *17*(3), 198–215. https://doi.org/10.1080/24732850.2017.1317560

Serin, R. C., Chadwick, N., & Lloyd, C. D. (2016). Dynamic risk and protective factors. *Psychology, Crime & Law*, *22*(1-2), 151–170. https://doi.org/10.1080/1068316X.2015.1112013

Serin, R. C., Gobeil, R., & Preston, D. L. (2009). Evaluation of the Persistently Violent Offender treatment program. *International Journal of Offender Therapy and Comparative Criminology*, *53*(1), 57–73. http://dx.doi.org/10.1177/0306624X07313985

Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and Violent Behavior*, *18*(1), 32–53. http://dx.doi.org/10.1016/j.avb.2012.09.002

Shah, S. A. (1978). Dangerousness: A paradigm for exploring some issues in law and psychology. *American Psychologist*, *33*(3), 224–238. https://doi.org/10.1037/0003-066X.33.3.224

Sheldon, K., & Gallagher, J. (2010). Risk reductions in dangerous and severe personality disordered forensic in-patients measured with the HCR-20 violence risk assessment scheme. *European Psychiatry*, *25*, 35. https://doi.org/10.1016/S0924-9338(10)70035-3

Shieh, G., Jan, S.-L., & Randles, R. H. (2007). Power and sample size determinations for the Wilcoxon signed-rank test. *Journal of Statistical Computation and Simulation*, *77*(8), 717–724. https://doi.org/10.1080/10629360600635245

Simourd, D. J. (2004). Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior*, *31*(3), 306–323. https://doi.org/10.1177/0093854803262507

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Pham, T. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, K., & Otto, R. K. (2014). International Perspectives on the Practical Application of Violence Risk Assessment: A Global Survey of 44 Countries. *International Journal of Forensic Mental Health*, *13*(3), 193–206. https://doi.org/10.1080/14999013.2014.922141

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice and Behavior*, *37*(9), 965–988. https://doi.org/10.1177/0093854810374274

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499–513. https://doi.org/10.1016/j.cpr.2010.11.009

Singh, J. P., Yang, S., Mulvey, E. P., & The RAGEE Group. (2015). Reporting guidance for violence risk assessment predictive validity studies: The RAGEE Statement. *Law and Human Behavior*, *39*(1), 15–22. https://doi.org/10.1037/lhb0000090

Skeem, J. (2013). Risk technology in sentencing: Testing the promises and perils (Commentary on Hannah-Moffat, 2011). *Justice Quarterly*, *30*(2), 297–303. https://doi.org/10.1080/07418825.2012.687513

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, *20*(1), 38–42. https://doi.org/10.1177/0963721410397271

Skeem, J. L., Monahan, J., & Mulvey, E. P. (2002). Psychopathy, treatment involvement, and subsequent violence among civil psychiatric patients. *Law and Human Behavior*, *26*(6), 577–603. https://doi.org/10.1023/A:1020993916404

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). Oxford University 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. https://doi.org/10.1023/A:1005513818748

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. https://doi.org/10.1037/0022-006X.63.6.1044

Spieker, S. J., Larson, N. C., Lewis, S. M., Keller, T. E., & Gilchrist, L. (1999). Developmental trajectories of disruptive behavior problems in preschool children of adolescent mothers. *Child Development*, *70*(2), 443–458. https://doi.org/10.1111/1467-8624.00032

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*(3), 265–271.

Steadman, H. J., & Cocozza, J. (1974). *Careers of the criminally insane*. Health.

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). Sage Publications.

Stratford, P. W., Binkley, J. M., Riddle, D. L. (1996). Health status measures: Strategies and analytic methods for assessing change scores. *Physical Therapy*, *76*(10), 1109–1123. https://doi.org/10.1093/ptj/76.10.1109

Tardiff, K. (1996). *Concise guide to assessment and management of violent patients* (2nd ed.). American Psychiatric Press.

Tarasoff v. Regents of California, 17 Cal. 3d 425:434 (1976).

Taxman, F. S., Caudy, M. S., & Pattavina, A. (2013). Risk-Need-Responsivity (RNR): Leading towards another generation of the model. In F. S. Taxman, & A. Pattavina (Eds.), Simulation strategies to reduce recidivism: Risk Need Responsivity (RNR) modeling in the criminal justice system (pp. 285–307). Springer.

Tengström, A., Hodgins, S., Müller-Isberner, R., Jöckel, D., Freese, R., Özokyay, K., & Sommer, J. (2006). Predicting violent and antisocial behavior in hospital using the HCR-20: The effect of diagnoses on predictive accuracy. *International Journal of Forensic Mental Health*, *5*(1), 39–53. https://doi.org/10.1080/14999013.2006.10471229

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment*, *14*(2), 139–153. https://doi.org/10.1177/107906320201400205

Thornton, D. (2016). Developing a theory of dynamic risk. *Psychology, Crime & Law*, *22*(1-2), 138–150. https://doi.org/10.1080/1068316X.2015.1109092

United States v. Salerno, 481 U.S. 739 (1987).

van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment*, *30*(2), 179–191. https://doi.org/10.1037/pas0000454

van den Berg, J. W., Smid, W., Kossakowski, J. J., van Beek, D., Borsboom, D., Janssen, E., & Gijs, L. (2020). The application of network analysis to dynamic risk factors in adult male sex offenders. *Clinical Psychological Science*, *8*(3), 539–554. https://doi.org/10.1177/2167702620901720

van der Put, C. E., Deković, M., Stams, G. J. J. M., van der Laan, P. H., Hoeve, M., & van Amelsfort, L. (2011). Changes in risk factors during adolescence: Implications for risk assessment. *Criminal Justice and Behavior*, *38*(3), 248–262. https://doi.org/10.1177/0093854810391757

Van Voorhis, P., & Presser, L. (2001). Classification of women offenders: A national assessment of current practices. US Department of Justice, National Institute of Corrections.

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, *37*(3), 261–288. https://doi.org/10.1177/0093854809357442

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. https://doi.org/10.1080/14999013.2012.737407

Viljoen, J. L., Gray, A. L., Shaffer, C., Bhanwer, A., Tafreshi, D., & Douglas, K. S. (2017). Does reassessment of risk improve predictions? A framework and examination of the SAVRY and YLS/CMI. *Psychological Assessment*, *29*(9), 1096–1110. https://doi.org/10.1037/pas0000402

Viljoen, J. L., Nicholls, T. L., Cruise, K. R., Desmarais, S. L., & Webster, C. D. (2014). *Short-Term Assessment of Risk and Treatability: Adolescent Version (START:AV) – User guide*. Mental Health, Law, and Policy Institute, Simon Fraser University.

Viljoen, J. L., Shaffer, C. S., Gray, A. L., & Douglas, K. S. (2017). Are adolescent risk assessment tools sensitive to change? A framework and examination of the SAVRY and the YLS/CMI. *Law and Human Behavior*, *41*(3), 244–257. https://doi.org/10.1037/lhb0000238

Vincent, G. M., Chapman, J., & Cook, N. (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. https://doi.org/10.1177/0093854810386000

Vitacco, M. J., Gonsalves, V., Tomony, J., Smith, B. E. R., & Lishner, D. A. (2012). Can standardized measures of risk predict inpatient violence?: Combining static and dynamic variables to improve accuracy. *Criminal Justice and Behavior*, *39*(5), 589–606. https://doi.org/10.1177/0093854812436786

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

Ward, T. (2014). The explanation of sexual offending: From single factor theories to integrative pluralism. *Journal of Sexual Aggression*, *20*(2), 130–141. https://doi.org/10.1080/13552600.2013.870242

Ward, T. (2015). Detection of dynamic risk factors and correctional practice: Criminogenic needs. *Criminology & Public Policy*, *14*(1), 105–111. https://doi.org/10.1111/1745-9133.12115

Ward, T. (2016). Dynamic risk factors: Scientific kinds or predictive constructs. *Psychology, Crime & Law*, *22*(1-2), 2–16. https://doi.org/10.1080/1068316X.2015.1109094

Ward, T., & Beech, A. R. (2015). Dynamic risk factors: A theoretical dead-end? *Psychology, Crime & Law*, *21*(2), 100–113. https://doi.org/10.1080/1068316X.2014.917854

Ward, T., & Fortune, C.-A. (2016). From dynamic risk factors to causal processes: A methodological framework. *Psychology, Crime & Law*, *22*(1-2), 190–202. https://doi.org/10.1080/1068316X.2015.1117080

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). Kluwer Academic.

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

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START;* version 1.1). British Columbia Mental Health and Addiction Services.

Webster, C. D., & Menzies, R. J. (1989). The clinical prediction of dangerousness. In D. N. Weisstaub (Ed.), *Coping with psychiatric and psychological testimony* (5th ed., pp. 22–38). Law and Psychology Press.

Webster, C. D., Nicholls, T. L., Martin, M., Desmarais, S. L., & Brink, J. (2006). Short-Term Assessment of Risk and Treatability (START): The case for a new structured professional judgment scheme. *Behavioral Sciences & the Law*, *24*(6), 747–766. https://doi.org/10.1002/bsl.737

Weems, K. (2007). Measurement error. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 1, p. 586). Sage Publications.

Whittington, R., Bjørngaard, J. H., Brown, A., Nathan, R., Noblett, S., & Quinn, B. (2014). Dynamic relationship between multiple START assessments and violent incidents over time: A prospective cohort study. *BMC Psychiatry*, *14*(1), 323. https://doi.org/10.1186/s12888-014-0323-7

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics, 1*, 80–83. https://doi.org/10.2307/3001968

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 Behaviour*, *37*(6), 377–388. https://doi.org/10.1037/lhb0000025

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. https://doi.org/10.1207/s15327752jpa8201_10

Woessner, G., & Schwedler, A. (2014). Correctional treatment of sexual and violent offenders: Therapeutic change, prison climate, and recidivism. *Criminal Justice and Behavior*, *41*(7), 862–879. https://doi.org/10.1177/0093854813520544

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. https://doi.org/10.1037/1076-8971.12.3.279

Wright, J. G., & Young, N. L. (1997). A comparison of different indices of responsiveness. *Journal of Clinical Epidemiology*, *50*(3), 239–246. https://doi.org/10.1016/S0895-4356(96)00373-3

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. https://doi.org/10.1037/a0020473

Appendix

Items on the HCR-20 and START

Table A-1. Items on the HCR-20 Risk Assessment Tool

Scale	Item
Historical (H) 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 (C) Scale	
C1	Lack of Insight
C2	Negative Attitudes
C3	Active Symptoms of Major Mental Illness
C4	Impulsivity
C5	Unresponsive to Treatment
Risk Management (R) 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 A-2. Items on the START Risk Assessment Tool

Vulnerability & Strength Scales	Item
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)