PERSONALITY TRAITS AND DYNAMIC VARIABLES ASSOCIATED WITH TYPES OF AGGRESSION IN HIGH SECURITY FORENSIC PSYCHIATRIC INPATIENTS.

by

Calvin M. Langton

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Human Development and Applied Psychology
Ontario Institute for Studies in Education
University of Toronto

© Copyright by Calvin M. Langton 2010
PERSONALITY TRAITS AND DYNAMIC VARIABLES ASSOCIATED WITH TYPES OF AGGRESSION IN HIGH SECURITY FORENSIC PSYCHIATRIC INPATIENTS.

Doctor of Philosophy 2010

Calvin Michael Langton

Human Development and Applied Psychology

University of Toronto

ABSTRACT

The Dangerous and Severe Personality Disorder (DSPD) initiative in England and Wales, underway since 2000, provides specialized care to high risk personality disordered individuals in prison and secure psychiatric facilities. Entry to the service, for a capacity nationwide total of approximately 300 individuals at four sites, is determined in part by risk (whether or not the individual is more likely than not to commit an offence that might be expected to lead to serious physical or psychological harm from which the victim would find it hard to recover). This requires valid procedures for assessing risk to determine individuals’ suitability for entry into and transfer out of the service. Yet little is known about the validity of current risk assessment tools and personality measures with the DSPD population. One of the studies reported, the first of its kind with the DSPD population, described a prospective evaluation of the predictive accuracy of the HCR-20, VRS, Static-99, and Risk Matrix 2000 with 44 admissions to the DSPD unit at a high security forensic psychiatric hospital. Consistent with hypotheses, all tools predicted damage to property. HCR-20 Total and scale scores predicted interpersonal physical
aggression with structured final risk judgments also predicting repetitive (2+ incidents of) interpersonal physical aggression. HCR-20 Risk Management scores were significantly associated with imminence of interpersonal physical aggression. The second study described a prospective evaluation of the predictive accuracy of Psychopathy Checklist-Revised Factor and Facet scores as well as scores for Cluster B traits using the International Personality Disorder Examination with the same sample. Partial support for hypotheses was found. Only Borderline PD dimension scores predicted damage to property. Histrionic PD predicted interpersonal physical aggression, and Histrionic, Borderline, and Antisocial PDs all predicted repetitive interpersonal physical aggression. Factor 1 and Facets 1 and 2 were also significant predictors of interpersonal physical aggression. Factor 1 and Histrionic PD scores were also significantly associated with imminence of this type of aggression. Results were discussed in terms of the practical utility of these tools with high risk forensic psychiatric inpatients and the functional link (between personality disorder and violence) criterion for DSPD service entry.
ACKNOWLEDGEMENTS

Among the many people due thanks for their support and assistance during the completion of this work, a number merit particular mention. I am very grateful to my thesis advisory committee members, Joe Ducharme, Judy Wiener, and Katreena Scott, for their willingness to serve and for their interest, patience, and feedback at crucial stages of the project. I owe my supervisor, Joe Ducharme, a particular debt of gratitude for his many years of support and encouragement. I am also very grateful to Michael Langton for his careful review of each of the chapters. Considerable thanks are due as well to my colleagues at the Peaks Unit at Rampton Hospital: Kevin Howells, Todd Hogue, Michael Daffern, and Aisling Mannion. Collaboration with these individuals has been a privilege. I would also like to acknowledge the work of the clinical staff at the Peaks Unit at Rampton Hospital, and the involvement of the patients. Without these two groups this project would not have been possible. Lastly, I would like to thank my wife, Sylvia. The really important things in my life, I have because of her.
TABLE OF CONTENTS

ABSTRACT .................................................................................................................. ii
ACKNOWLEDGEMENTS .......................................................................................... iv
TABLE OF CONTENTS ............................................................................................ v
LIST OF TABLES ....................................................................................................... viii
LIST OF FIGURES .................................................................................................... ix
LIST OF APPENDICES ............................................................................................. x
CHAPTER 1: GENERAL INTRODUCTORY COMMENTS ......................................... 1
CHAPTER 2: STUDY #1 ............................................................................................. 10
  2.1 INTRODUCTION ............................................................................................... 11
  2.2 METHOD .......................................................................................................... 18
    2.2.1 Setting and Participants ........................................................................... 18
    2.2.2 Measures .................................................................................................. 19
    2.2.3 Procedure ................................................................................................ 24
    2.2.4 Data Analysis ........................................................................................... 26
  2.3 RESULTS .......................................................................................................... 27
    2.3.1 Relationships Between the Assessment Tools ....................................... 28
    2.3.2 Predicting Aggressive Behaviour ............................................................ 32
    2.3.3 Predicting Repetitive Aggressive Behaviour ........................................... 38
    2.3.4 Imminence of Outcomes ......................................................................... 39
  2.4 DISCUSSION ..................................................................................................... 42
    2.4.1 Prediction Accuracy .................................................................................. 43
    2.4.2 Relationships Between Measures ............................................................. 49
LIST OF TABLES

CHAPTER 2

Table 2.1. Correlations Between Component Scores on the HCR-20, VRS, and PCL-R..........................................................30

Table 2.2. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the Risk Assessment Tools for Incidents Involving Damage to Property .............33

Table 2.3. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the Risk Assessment Tools for Incidents Involving Interpersonal Physical Aggression.................................................................35

CHAPTER 3

Table 3.1. Correlations Between Measures.................................................65

Table 3.2. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the PCL-R and IPDE Dimensions for Incidents Involving Damage to Property ..........................................................67

Table 3.3. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the PCL-R and IPDE Dimensions for Incidents Involving Interpersonal Physical Aggression.................................................................69
LIST OF FIGURES

CHAPTER 2

Figure 2.1. Interpersonal physical aggression survival curves for HCR-20 Risk Management (RM) quartile groups……………………………41

CHAPTER 3

Figure 3.1. Interpersonal physical aggression survival curves for PCL-R Factor 1 quartile groups……………………………………………..75

Figure 3.2. Interpersonal physical aggression survival curves for Histrionic PD quartile groups……………………………………………76
LIST OF APPENDICES

Appendix 1. Glossary of Terms.................................................................134
Appendix 2. Participant Information Sheet.............................................135
Appendix 3. Consent Form.................................................................138
CHAPTER 1: General Introductory Comments

Inpatient aggression represents a significant concern and imposes considerable financial costs for agencies as well as psychological costs for service providers and fellow patients (Carmel & Hunter, 1993; Hunter & Carmel, 1992; Daffern, Mayer, & Martin, 2004; Martin & Daffern 2006). Clinically, the challenge is to balance the provision of care for mental health patients with the management of risk for violence against staff and fellow patients that they might pose. Failure to ensure staff and patient safety undermines intervention efforts and compromises the therapeutic environment. Being able to identify those who threaten the safety of others is therefore a high priority both for maintaining safety and for the effective provision of services.

A substantial amount of the research on violence potential has focused on long-term chronic risk, that is, risk for violence over a period of many months or years (Monahan et al., 2001), but there is a growing body of research that is concerned with risk for violence in the short-term, that is, acute risk over a period of hours, days, or weeks (McNeil, 2009). The distinction is meaningful because sets of variables have been shown to vary in terms of their predictive validity for events such as a type of violence or criminal behaviour according to their temporal stability and the timeframe over which the prediction is assumed to be relevant. Two general categories of predictors are static and dynamic risk factors. Static factors are characteristics of an individual or her or his situation that are historical in nature (such as childhood maladjustment, criminal record, previous diagnoses) or are at least stable over time (such as temperament or personality traits), and are generally immune to planned interventions provided in the short term). To function as a risk factor, they must temporally precede the adverse event to be predicted
(e.g., inpatient violence), often being quite distal in time from the event, and increase its likelihood of occurring.

In contrast, dynamic factors are characteristics of an individual or her or his situation that are part of the presentation at the time of assessment (such as current diagnoses, symptoms of mental illness, compliance with treatment, coping or social skills deficits, social supports, criminal associates, and access to victims, weapons, or drugs). To function as predictors they too must precede the adverse event, generally featuring more proximally in time, but must also be capable of changing (whether spontaneously, through maturational processes, or due to intervention) and the change must be associated with a change in the likelihood of the event occurring (Douglas & Skeem, 2005).

A number of risk assessment instruments with demonstrated predictive validity are composed entirely of static factors and scores on these are associated with probabilities of engaging in future violence over a period of years (e.g., Hanson & Thornton, 2000; Harris, Rice, & Quinsey, 1993). In contrast, dynamic variables appear to be superior for predicting adverse events such as violence over much shorter time periods (McNeil, 2009) and a variety of tools have been developed and validated for this purpose (e.g., Almvik, Woods, & Rasmussen, 2000; McNeil, Gregory, Lam, Binder, & Sullivan, 2003; Nicholls, Brink, Desmarais, Webster, & Martin, 2006; Ogloff & Daffern, 2006).

When selecting a risk assessment tool, a number of factors require consideration. Most pertinent for the studies reported in the following chapters are questions concerning population, outcome, and time frame. For use with which population was the tool developed? Can the individual to be assessed reasonably be considered a member of that
population (for example, forensic or civil psychiatric patients, general prison/correctional population, sexual offenders or perpetrators of domestic violence)? If the individual to be assessed is arguably different in some way from the population for which the tool was originally intended, is there empirical evidence to support its use with the population that the patient more closely resembles? Has the tool been shown to predict the adverse event of concern and for which risk management strategies need to be devised (e.g., general criminal recidivism, violence in general, institutional violence, violence towards a spouse, or sexual violence)? Over what time frame has the tool been empirically demonstrated to predict the adverse event (e.g., days, weeks, months, or years) and for what time frame is a prediction required from the assessment? If repeatedly administered, will the tool capture changes in risk over time (i.e., does it include dynamic risk factors) or does it serve to classify the patient in terms of long-term risk (on the basis of static factors)? What other assessment protocols (i.e., measures of constructs such as personality traits) are applicable to the population to which this patient belongs and can these protocols inform risk assessment and management efforts as well?

The two studies that are described in the following chapters represent an attempt to evaluate the predictive validity of the various personality measures and risk assessment tools (some of which incorporate dynamic risk factors) included within the core battery used with patients admitted into the Dangerous and Severe Personality Disorder (DSPD) service in the United Kingdom (DSPD Programme, Department of Health, Ministry of Justice, and HM Prison Service, 2008). The service is designed to provide specialized care to high risk offenders with mental disorders. These offenders have been identified from the mental health and prison services on the grounds that they have a severe
disorder of personality which renders them a high risk of causing serious physical or psychological harm to others. Of course, not everyone with a personality disorder engages in criminal or physically violent behaviour and not all personality disordered offenders would meet criteria for entry into the DSPD service. Prevalence rate for personality disorder in the general population in Great Britain is estimated to be 4.4% (Coid, Yang, Tyrer, Roberts, & Ullrich, 2006) while among males prisoners on remand and sentenced in England and Wales the rates are 78% and 64%, respectively, out of total populations of 12,302 and 46,872, respectively (Singleton, Meltzer, Gatward, Coid, & Deasy, 1998). Although estimates at the time the proposal for the DSPD service was published put the number of males 18 years or older who would meet criteria for the service well in excess of 2,000 (1,400 in prison, 400 detained on court orders in secure psychiatric hospitals, and 300 to 600 not currently detained but frequently involved in the legal and mental health systems; Home Office and Department of Health, 1999), the ongoing pilot phase of the initiative involves only 300 beds in four high security sites (two in prison settings, two in psychiatric hospital settings) across England and Wales (DSPD Programme et al., 2008).

The DSPD initiative was launched as part of an effort by the British Government to improve the protection of the public from dangerous and severely personality disordered individuals. Two components underscored this general objective: to ensure that dangerous and severely personality disordered individuals are detained in a secure setting until the risk they pose is no longer high (estimated to require a minimum of three years; National Probation Service, 2008), and to manage them in a manner that ensures appropriate services to address their disorders. The aims at the outset were to develop a
service that “strikes the right balance between the interests of individuals and of society; meets the needs of this group of people better than the present patchy provision; is firmly grounded in evidence from research, and capable of adapting over time as new research evidence comes forward; provides better value for money than the present arrangements; leads in time to a reduction in the level of the most serious offending by people with severe personality disorder, as better preventive measures are identified and implemented, and through the early identification and detention of those who are dangerous.” (Home Office and Department of Health, 1999, pp. 4-5). Dedicated sites with specialized programs and specially trained staff were deemed necessary to meet these aims. The applicable legislation is found in the Mental Health Acts 1983 and 2007 and the Criminal Justice Act 2003 (Bickle, 2008). Importantly, individuals admitted to the DSPD units have not necessarily been referred on a voluntary basis and so developing motivation and engagement are central to the assessment and treatment process (DSPD Programme et al., 2008; cf. Parhar, Wormith, Derksen, & Beauregard, 2008).

Punishment, treatment, and risk are explicitly seen as overlapping concepts within the initiative: “the risk an individual represents may well only come to light after they have committed a criminal act for which they should be punished… [and their] risk may sometimes be reduced through treatment.” (Home Office and Department of Health, 1999, p. 9).

It is worth briefly noting that attempts enshrined in law to (i) protect the public from individuals deemed dangerous on the basis of their abnormal behaviours, and to (ii) meet the needs of these individuals have a well-documented history in the UK and elsewhere and have given rise to a variety of legal, ethical, and clinical concerns (see, for
example, Bonta, Zinger, Harris, & Carrière, 1998; Jakimiec, Porporino, Addario, & Webster, 1986; Johnson, 2004; Maden, 2007; Peay, 2007; Petrunik, 1994; Pratt, 1996; Wormith & Ruhl, 1987). Although the scope of legislation enabling indeterminate detention of individuals across countries is highly variable, most involve reviews within set time periods and reflect either a community protection model or a clinical model (Home Office and Department of Health, 1999). The former emphasizes public safety over the civil rights of the individual, with legal powers of indefinite detention on the basis of the risk posed. The latter emphasizes diagnosis and treatment. The DSPD initiative has explicitly attempted to reflect both, but has not been without controversy (Gunn, 2000; Haddock, Snowden, Dolan, Parker, & Rees, 2001; Moran, 2001). One aspect that has caused much concern is the centrality of risk in the initiative but also more broadly as a potential driving force in contemporary mental health and penal policy (Corbett & Westwood, 2005; cf. Seddon, 2008). It is with an examination of risk, specifically, and the implications of its assessment for clinical decision-making and intervention work that the studies in the following chapters are concerned.

Certainly, the DSPD program does place considerable emphasis on risk status as part of the entry criteria for the service. In practice, the assessment battery prescribed for use with the DSPD population provides mental health professionals and direct care providers with a broad overview of the psychiatric and psychosocial needs and personality profiles of each patient. But quite apart from supplemental assessment protocols related to specific aspects of a patient’s clinical presentation or planned treatment initiatives, the core battery’s potential utility lies perhaps most clearly in terms of the picture it generates of the clinical features associated with an increased risk of
patient violence as well as certain personality traits. These personality traits will require intervention, of course, but they will also impact on the content and format in which all other interventions (for Axis I disorders and criminogenic needs, for example) are planned and delivered. Viewed principally from this perspective, the battery should serve as the foundation for patients’ individualized treatment and risk management plans. Yet little research has emerged concerning the predictive validity and clinical utility with the DSPD population of the key tools comprising this battery (cf. Duggan, Mason, Banerjee, & Milton, 2007). It was to begin addressing this gap in the knowledge base that the evaluation project to be described was undertaken.

The main research question was whether four of the risk assessment tools included in the battery (the Historical Clinical Risk Management 20 [HCR-20: Webster, Douglas, Eaves, & Hart, 1997], the Violence Risk Scale [VRS: Wong & Gordon, 2000], the Risk Matrix 2000 scales [Thornton et al., 2003], and the Static-99 [Hanson & Thornton, 2000]) and the two measures of traits associated with psychopathy and personality disorders (the Psychopathy Checklist-Revised [PCL-R: Hare, 2003] and the International Personality Disorder Examination [IPDE: Loranger et al., 1994]) accurately predicted two types of institutional aggression exhibited by personality disordered patients in a high security forensic psychiatric hospital using a prospective design. Given the empirical evidence (reviewed in detail in the following chapters) demonstrating predictive validity for some of these tools and at least a statistical association with types of aggression for others, it was hypothesized that all would significantly predict acts involving damage to property and interpersonal physical aggression committed by the initial set of consecutive admissions to one of four national DSPD sites.
Of course, there are distinct literatures associated with personality disorder and violence risk assessment (although, as one might expect, there is some overlap, notably in the area of psychopathy). As such, presentation of the results of this evaluation project is divided into two chapters, affording a chance to consider the most pertinent literature in each area in some depth while drawing links between the two where appropriate.

In the next chapter, attention is focused on the four risk assessment tools. The empirical evidence for the predictive validity of each is reviewed, concentrating particularly on the extant research with forensic psychiatric inpatients. It will be noted that there are considerable differences in the scope and depth of this evidence base for each, even though all have been proscribed for use with the DSPD population, and there are sufficient grounds to hypothesize that all predict types of inpatient aggression. The reasoning for (re-)evaluating the validity of a given tool for distinct populations and various settings is explored. The study design and data analytic strategy for the project is then described, noting in particular the importance of controlling for variable follow-up periods within and between samples as well as looking at imminence of aggression as a dimension of predictive validity. Following presentation of the results, the discussion explores the implications of the findings in light of the methodological limitations of the study as well as the potential utility of dynamic risk factors with inpatients. Also considered are the associations between component scales of the tools and the construct of psychopathy, the predictive validity of which is itself examined alongside other traits of personality disorders in the second study to be reported.

In the third chapter, the emphasis shifts from atheoretical purpose-designed risk assessment tools to measures of personality traits associated with psychopathy and
personality disorders. The burgeoning literature on the predictive validity of the PCL-R is seen to contrast sharply with the relative dearth of studies exploring the association between IPDE Cluster B dimension scores and aggression. Importantly, however, there is sufficient evidence to postulate for each the accurate prediction of inpatient acts of aggression in the present sample. Following a review of the study design and data analytic strategy, the results are presented, again including imminence as a dimension of predictive validity worthy of attention and looking also at the strength of associations between predictors and outcomes while controlling for additional variables. A main consideration in the discussion concerns methodological issues that might explain some of the consistencies and inconsistencies between findings reported here for the Factors and Facets of the PCL-R and those found by other researchers. Another is the clinical implications of Cluster B and psychopathic traits for treatment planning, both in terms of criminogenic needs to be targeted and specific responsivity factors informing delivery of services.

In the final chapter, following an overview of the project objectives and main findings, limitations of the work are revisited. Recent conceptual developments and their implications for understanding differences in the extant research are introduced. Attention then shifts to more general implications and future avenues of investigation, particularly those that would integrate the current DSPD assessment battery within a more systemic approach tied theoretically and clinically with service delivery and treatment evaluation work.
CHAPTER 2: Study #1

Prediction of institutional aggression among personality disordered forensic patients using actuarial and structured clinical risk assessment tools: Prospective evaluation of the HCR-20, VRS, Static-99, and Risk Matrix 2000.1

Abstract

Entry to the Dangerous and Severe Personality Disorder (DSPD) service in England and Wales is heavily determined by risk status, and therefore requires valid procedures for monitoring changes in risk over time in order to make risk management decisions and determine patients’ suitability for transfer to lower security settings. Yet little is known about the validity of current risk assessment tools with the new DSPD population. This study reports a prospective evaluation of the predictive accuracy of the HCR-20, VRS, Static-99, and Risk Matrix 2000 with 44 admissions to the DSPD unit at a high security forensic psychiatric hospital. Thirty nine percent of the sample exhibited interpersonal physical aggression on one or more instances over an average 1.5 year period following admission, and a similar percentage caused damage to property on one or more occasions over the same period. All tools predicted damage to property. HCR-20 Total and scale scores predicted interpersonal physical aggression with structured final risk judgments also predicting repetitive (2+ incidents of) interpersonal physical aggression. HCR-20 Risk Management scores were significantly associated with imminence of interpersonal physical aggression. Results were discussed in terms of the practical utility of these tools with high risk forensic psychiatric inpatients.

2.1 Introduction

There has been considerable progress in the field of violence prediction and risk assessment procedures in recent years, marked in part by the emergence of a large array of tools with demonstrated predictive accuracy for various psychiatric and correctional populations (Daffern, 2007; Monahan et al., 2001; Hanson & Morton-Bourgon, 2007). Unfortunately, clinicians’ ability to determine when an individual’s risk changes, whether through treatment, maturation, or otherwise, has not kept pace with the improved ability of clinicians to rank individuals in terms of their likelihood of committing future violent or criminal acts using these tools (Becker & Murphy, 1998; Douglas & Skeem, 2005; Hanson, 1998).

In regard to the risk assessment tools most widely used in forensic work, this limitation is due in part to the almost exclusive reliance on static variables (e.g., criminal history) as predictors. These tools represent what Bonta (1996) referred to as second generation empirically based assessments, which have advanced the field by improving upon first generation approaches reflecting professional judgment alone. It is with the emergence of third generation, evidence-based dynamic risk assessment tools that clinicians should be able to monitor intra-individual variation in risk over time (Bonta & Andrews, 2007). For now, however, the evidence remains limited in terms of whether current tools are sensitive to change in risk and whether such change is empirically associated with a change in likelihood of particularly outcomes such as violence (see Raynor, 2007, and Quinsey, Jones, Book, & Barr, 2006, for encouraging data in this regard).
The Dangerous and Severe Personality Disorder (DSPD) initiative in England and Wales (DSPD Programme et al., 2008) is a program for which entry is heavily determined by risk status (risk for future violence), and which therefore requires valid procedures for monitoring changes in risk over time in order to make risk management decisions and determine patients’ suitability for transfer to lower security settings. The DSPD programme was implemented to provide specialized care to high risk personality disordered offenders (see Howells, Langton, & Hogue, 2007, and Tyrer, 2007, for collected papers). Little research has yet emerged concerning the predictive validity and clinical utility with the DSPD population of the key tools prescribed for use in DSPD settings. Coid et al. (2007) recently reported a community follow-up study of 1,396 adult male offenders in the UK who were at risk for reoffending over a mean time of 1.97 years post-release. Among the tools evaluated, Coid et al. included the Historical Clinical Risk-20 (HCR-20: Webster et al., 1997), the Risk Matrix 2000 Violence scale (RM 2000/V: Thornton et al., 2003), and the Psychopathy Checklist-Revised (PCL-R: Hare, 2003). All three were found to significantly predict criminal recidivism, including violence specifically, with Areas under the Receiver Operating Characteristic (AUCs) ranging from .59 to .69. Of particular interest, Coid et al. found that 15% of the cohort would have met criteria for entry into the DSPD service, with these offenders receiving significantly more reconvictions, notably for violent offences, over the follow-up period.

2 Although a number of indices of predictive validity are reported in the literature, the area under the Receiver Operating Characteristic (ROC area or AUC) is the accepted measure of predictive and diagnostic accuracy (Swets, Dawes, & Monahan, 2000), representing an index of the trade-off between sensitivity and specificity as a function of test score (Harris, Rice, & Quinsey, 2007). There is not, however, agreement regarding an accepted standard for interpreting AUC values. Rice and Harris (2005) provide a table (pp. 616-617) for comparing the most common current measures of effect sizes. They report that an AUC value of .556 equates to a Cohen’s d of .200, considered a small effect size, an AUC value of .639 equates to a Cohen’s d of .500, considered a medium effect, and an AUC value of .714 equates to a Cohen’s d of .800, considered a large effect size.
compared with the comparison group. However, Coid et al. did not examine the predictive accuracy of the tools with just the subset meeting DSPD entry criteria nor did they report data regarding repeated assessments of the dynamic scales of the HCR-20 (the Clinical and Risk Management scales).

The present study is part of a project intended to address some of the gaps in what is known about the utility of the main tools recommended for use with the DSPD population (DSPD Programme et al., 2008). One important question is whether or not these measures actually predict types of institutional infractions, particularly physically aggressive behaviour, among personality disordered patients in a high security forensic psychiatric setting. The predictive validity of the (majority of the) risk assessment tools recommended for use within the DSPD programme has been established for use with certain populations and settings (Coid et al., 2007; Hanson & Morton-Bourgon, 2007; Leistico, Salekin, DeCoster, & Rogers, 2008) but it is important to understand what is being determined in terms of a tool’s predictive validity when one moves from community-based settings to an institutional context (or from high security to lower security, cf. Dernevik, Grann, & Johansson, 2002), and from relatively large samples, often of convenience (e.g., all offenders released from a particular institution in a given period), to a selected subset. The DSPD population, in theory, consists of those individuals from the prison and mental health systems that pose very high risk of committing future acts of violence. The tools have been used to identify this subset from the population of possible candidates. But the same measures may not accurately predict violent behaviour in only this subset (because, for example, the range of scores is now severely attenuated) or in this setting (because, for example, the opportunities and
antecedents associated with violence in community settings might be quite different in function or form to those in a high security institutional setting; Daffern, Ferguson, Ogloff, Thomson, & Howells, 2007). So the question is whether these tools are useful for appraising and managing ongoing risk of violence for individuals transferred, on the basis of their high scores on these tools, into the DSPD programme.

For the risk assessment tools that supplement demographic and criminal history variables with items tapping clinical and contextual concerns, whether they are sensitive to change in risk level over time represents another important question. To address this, however, it is necessary to have first demonstrated predictive validity in the context or setting in which ongoing risk management decisions will be undertaken. This is the focus of the present study.

Evidence from the wider criminological literature shows that treatment-related reductions in institutional infractions are associated with lower rates of post-discharge recidivism (French & Gendreau, 2006). Given this, interventions during participation in the DSPD service that demonstrably reduce the likelihood of institutional infractions have important implications beyond the immediate concerns of patient and staff safety and security within the institution. Indeed, a reduction in institutional infractions could be equated with a reduction of risk for recidivism post-discharge (Wong et al., 2005). This suggests that being able to detect changes in risk for institutional infractions using a risk assessment tool would allow clinicians to make reasonable inferences about risk of recidivism post-discharge. However, if a measure does not predict adverse events within the institutional setting, the implications of a change in score on the measure are less straightforward. One could consider the individual whose risk score has decreased by
referring back to the population from which she or he was originally drawn as a DSPD candidate to determine if the lowered risk score now falls below the threshold used for selection, which would mean she or he no longer warrants inclusion in the DSPD subset. More immediately, however, if a measure does not predict adverse events within the institutional setting then a change in score on the measure provides only a weak basis for transfer to lower security settings because it indicates little about the current likelihood of adverse events occurring in such a setting (Duggan, 2007; Langton, 2007).

The DSPD assessment battery includes the HCR-20 (Webster et al., 1997), the Violence Risk Scale (VRS: Wong & Gordon, 2000), the RM 2000 (Thornton et al., 2003), the Static-99 (Hanson & Thornton, 2000) and the PCL-R (Hare, 2003). Among the risk assessment tools, the HCR-20 and the VRS are of particular interest because both reflect a risk-needs approach (Andrews & Bonta, 2006; Andrews, Bonta, & Wormith, 2006); that is, both include items (comprising the Clinical and Risk Management scales of the HCR-20 and the Dynamic Factors scale on the VRS) with a specific set of characteristics. These items are focused on features of the individual and her or his situation that can change over time, can be targeted in treatment, and for which observed changes (are hypothesised to) correspond to a change in the likelihood of recidivism.

While the HCR-20 has been subject to a relatively large number of evaluations (Campbell, French, & Gendreau, 2009; Douglas, Guy, & Weir, 2006), independent studies of the VRS have only more recently begun to appear in the peer-reviewed literature (Daffern, 2007; Wong & Gordon, 2006). Focusing here on those studies with forensic psychiatric inpatient samples, the HCR-20 and VRS have been shown to predict short-term (that is, over months rather than years) inpatient violence although evidence
for the VRS has been mixed. For example, over an initial 3-month period following admission to one of two medium-security forensic psychiatric facilities in the UK, Gray et al. (2003) found that scores on an abbreviated version of the HCR-20 (and of particular note, the Clinical scale) predicted verbal aggression (AUCs .73-.79), violence to property (AUCs .77-.83), and physical aggression (AUCs .77-.81) in their sample of 34 mentally disordered offenders (15% of whom had a personality disorder diagnosis). Similarly, Grevatt, Thomas-Peter, and Hughes (2004) reported that scores on the HCR-20 Clinical scale significantly predicted incidents of verbal abuse (AUC .81) and damage to property (AUC .65), but not physical assaults specifically (AUC .60), over a 6-month period in their UK sample of 44 patients held in an independent forensic psychiatric facility (39% of whom had a personality disorder). In the same study, however, Grevatt et al. found that the Historical scale failed to predict these outcomes and neither the VRS Total nor Static and Dynamic Factor scale scores were significantly predictive. In contrast, Dolan and Fullam (2007) found that the HCR-20 (Total, Historical, Clinical, and Risk Management scale scores) significantly predicted violence (AUCs .66-.73) as did the VRS (Total and Dynamic Factors scale scores, AUCs of .72 and .72) in a sample of 80 mentally disordered offenders over 12-months following admission to a medium security unit in the UK.

The two additional risk assessment tools included in the DSPD assessment battery and evaluated in the present study, the Static-99 (Hanson & Thornton, 2000) and the RM 2000 (Thornton et al., 2003), were both designed to facilitate the assessment of risk for criminal recidivism among sexual offenders. Importantly, neither of these tools incorporate dynamic risk factors, being comprised instead of static items scored using
offence history information. While neither can therefore provide indices of risk that could be sensitive to decreases over time, their inclusion in the DSPD battery is presumably justified on the grounds that both provide valid rankings of risk for future violence (upon release to community settings), which is required for determining DSPD service entry criteria. The developers of both have reported statistically significant levels of predictive accuracy for any violent (AUCs .69-.81) and sexual reoffending (AUCs .71-.77) using large samples of sexual offenders (drawn from forensic psychiatric as well as correctional settings in the case of the Static-99) followed-up over a period of years. Although other researchers have also reported data confirming the predictive validity of both with sexual offenders followed up after release (e.g., Craig, Beech, & Browne, 2006; Langton, Barbaree, Seto et al., 2007), to date, there are no studies in the peer-reviewed literature that have tested the ability of these scales to predict institutional misconduct or physical aggression in forensic psychiatric inpatient settings (R. K. Hanson, personal communication, October 2, 2007; D. Thornton, personal communication, September 29, 2007). This is a gap in the knowledge base that requires redress if either are to be given weight within the DSPD settings beyond entry.

The PCL-R is another measure included in the DSPD assessment battery. Although developed to assess the constellation of interpersonal, affective, and behavioural features associated with psychopathy rather than as a tool for assessing risk for future criminality or violence (Hare, 1996), the PCL-R has been found to be a relatively robust predictor of such behaviours in community and institutional settings with both general offenders and psychiatric samples. For example, in their meta-analytic investigation involving 95 studies with non-overlapping samples, Leistico et al. (2008)
reported medium effect sizes, with Hedges’ $d$ (mean weighted) values of 0.55, 0.38, and 0.60 for PCL Total, Factor 1 and Factor 2 scores, respectively. The concept of psychopathy is central to the DSPD initiative (DSPD Programme et al., 2008; Maden, 2007; Mullen, 2007). One reason for this is because it provides some theoretical and empirical grounds for explicating the required link between personality disorder and risk of causing harm to others. Evaluation of the predictive validity of the PCL-R with the current sample is reported in the next chapter in a companion study concentrating on measures of personality traits and disorders. Its inclusion in the current study was for the purposes of investigating the association between scores on the purpose-specific risk assessment tools and facets of psychopathy as operationalized by the PCL-R. Based on the extant literature, it was expected that the risk assessment tools would all predict institutional infractions among the study sample. Given the atheoretical nature of the risk assessment tools, no a priori hypotheses were formulated regarding associations between tool scores and scores on the PCL-R Facets.

2.2 Method

2.2.1 Setting and Participants

The Peaks Unit is a purpose-built inpatient facility intended for individuals detained under the DSPD initiative located within the perimeter of a high security hospital (see Hogue, Jones, Talkes, & Tennant, 2007, for description of philosophy and clinical services in the unit). Of the first 51 consecutive admissions to the Peaks Unit, assessment and incident data were available for 44 of these patients, all male. The mean age was 34.41 years ($SD = 8.47$, range = 18.55 to 54.44). The majority were Caucasian. Using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-CV: First,
Spitzer, Gibbon, & Williams, 1996), the most common diagnoses were for alcohol and substance-related disorders (57% and 52%, respectively), post-traumatic stress disorder (36%), and major depressive disorder (24%). Using the International Personality Disorder Examination (IPDE: Loranger et al., 1994), the most common personality disorder (PD) diagnoses were for Antisocial PD (73%), Borderline PD (41%), and Narcissistic PD (16%). Using the recommended UK cut-off score of 28 on the PCL-R (Cooke, Michie, Hart, & Clark, 2005), 49% of the sample met criteria for psychopathy: prorated PCL-R total scores ranged from 18.9 to 35, with a mean of 27.6 (standard deviation of 4.2). All patients had been charged with or convicted of a criminal offence, and 89% had committed one or more violent offences. Analyses using the full follow-up period involved these 44 participants. Analyses using an initial 12-month fixed follow-up period, utilized in order to control for unequal follow-up times between participants (see 2.2.3 Procedure), involved a subset of 36.

2.2.2 Measures

*Historical Clinical Risk – 20.*

The HCR-20 (Webster et al., 1997) was developed to structure the clinical assessment of risk for violence and to guide treatment planning with offender and psychiatric patient populations. It is comprised of 20 items, 10 *historical* in nature (e.g., static risk factors such as history of violence, mental illness, early maladjustment, relationship instability, and employment problems); 5 *clinical* in nature (e.g., dynamic risk factors such as active symptoms of mental illness, lack of insight, response to treatment); and 5 *risk management* items, which involve anticipation of risk-related factors (e.g., dynamic risk factors such as feasibility of plans, level of personal support,
non-compliance). Evaluators use clinical judgment to rate each of the items on a 3-point scale (0 = the item is definitely absent or does not apply, 1 = the item is possibly present, or is present to only a limited extent, 2 = the item is definitely present).

Items can be summed to give Total and scale scores. Total score can range from 0 to 40 with higher scores indicating higher risk of future violence. After rating the 20 items, the assessor can also make a structured final risk judgment (SFRJ), a 3-category ranking of Low, Medium or High, based on the structured professional judgment model, regarding risk for violence. In the present study, Risk Management items were coded for risk of physical aggression while institutionalised (i.e., as R-in assessments using a projected 4 to 6-month time frame, rather than as R-out assessments for which the R items are coded for risk of violence in the community post-release – see manual, p. 23). Scores were prorated on the basis of the number of omitted items as per the recommendation in the manual (p. 21). Douglas, Ogloff, and Hart (2003) have reported acceptable levels of inter-rater reliability for the HCR-20 Total and scale scores (intraclass correlation coefficients ranged from .47 to .94).

*Violence Risk Scale.*

The VRS (Wong & Gordon, 2000, 2006) is a risk assessment tool designed to facilitate violence assessment, prediction, and treatment planning. It was developed for

---

3 Items in individual cases were omitted when insufficient information to score the item was available or conflicting information could not be reconciled. No adjustment was made to scores on the Static-99 or RM 2000 scales when items were omitted. PCL-R Total, Factor, and Facet scores were prorated when items were omitted according the manual (Hare, 2003) as were the VRS Total, Static, and Dynamic Factor scale scores (Wong & Gordon, 2000). HCR-20 Total, Historical, Clinical, and Risk Management scale scores were prorated when items were omitted as per the general recommendation in the manual (Webster et al., 1997). In the absence of specific instructions, the procedure used was the same as that used for the VRS:

\[
\text{Prorated Score} = \frac{x \times y}{y - z}
\]

Where \(x\) is the total or scale score, \(y\) is the number of items used to calculate the total or scale score, and \(z\) is the number of omitted items.
use with correctional and forensic psychiatric populations. The most recent version of the VRS consists of 6 static risk factors (e.g., age at first violent conviction, prior release failures/escapes, stability of family upbringing) and 20 dynamic risk factors (e.g., criminal attitudes, insight into cause of violence, substance use, cognitive distortion). Evaluators use clinical judgment to rate the items on a 4-point scale (from 0 = characteristics described in the item do not apply to 3 = characteristics described in the item match those of the individual). Ratings are summed to provide scores for the Static Factors and Dynamic Factors scales as well as a Total score (range 0 to 78). Scores are prorated on the basis of the number of omitted items according to procedures detailed in the manual (p. 18). The dynamic items can be coded at pre- and post-treatment stages. In the present study only the initial intake VRS assessments (pre-treatment within the Peaks Unit) were used. Wong and Gordon (2006) reported high inter-rater reliability indices for the VRS (intraclass correlation coefficients ≥ .92).

Static-99.

The Static-99 (Hanson & Thornton, 2000; A. Harris, Phenix, Hanson, & Thornton, 2003) is an actuarial tool developed to assess risk for sexual and violent recidivism among adult males known to have committed at least one sexual offence. It has 10 items: number of prior charges or convictions for sexual offences; age upon release from prison or anticipated exposure to risk for reoffending in the community; any male victims (coded as yes or no); and any unrelated victims (coded as yes or no); number of prior sentencing dates; any convictions for non-contact sexual offences (coded as yes or no); non-sexually violent index offence dealt with at sentencing for index sex offence (coded as yes or no); prior non-sexually violent offence (coded as yes or no); any
stranger victims (coded as yes or no); and cohabitation status (ever lived with a lover for at least two years, coded as yes or no). Total scores can range from 0 to 12; individuals can be assigned to one of four risk levels (0-1 = Low, 2-3 = Medium-Low, 4-5 = Medium-High, 6+ = High) or seven risk levels (0 = lowest, 6+ = highest) based on their score. Tabulated indices of predictive accuracy in the present study are for the version with 7-risk levels or ‘bins’. Indices of inter-rater reliability reported elsewhere for the Static-99 have been acceptable (r = .90, Looman, 2006).

The Static-99 was included in the present study because it is completed for individuals with a history of sexual offences as part of the initial DSPD assessment battery. However, its exclusive reliance on static predictors meant it could not inform the more central interest here in the utility of dynamic variables and so only a subset of the analyses were carried with it. The same reasoning informed the decision to restrict the number of analyses using the Risk Matrix 2000 scales.

**Risk Matrix 2000 Scales.**

The 7-item (RM 2000/Sexual) and 3-item (RM 2000/Violence) scales (Thornton, 2005; Thornton et al., 2003) were designed to assess likelihood of sexual and violent recidivism, respectively, as distinct outcomes in offenders with a history of one or more sexual offences. The 7-item RM 2000/S scale is scored in two steps. In the first step, three items (age at date of release/start of at-risk period; number of occasions sentenced for a sexual offence; number of occasions sentenced for any criminal offence) are scored and summed to produce one of four risk categories: 0 = Low (numerical value of 1 for analyses), 1-2 = Medium (numerical value of 2 for analyses), 3-4 = High (numerical value of 3 for analyses), and 5-6 = Very High (numerical value of 4 for analyses). The
second step involves scoring four items considered aggravating factors (male victim of sexual offence; stranger victim of sexual offence, single/never married; non-contact sex offence). These items are scored and summed. Step 2 scores of 0 or 1 produce no change in the Step 1 risk category. Step 2 scores of 2 or 3 elevate the Step 1 risk category by one. A step 2 score of 4 elevates the Step 1 risk category by two.

The three items comprising the RM 2000/V (age at date of release/start of at-risk period; number of occasions sentenced for nonsexual violence; ever convicted for burglary) are scored and summed to produce one of four risk categories: 0-1 = Low (numerical value of 1 for analyses), 2-3 = Medium (numerical value of 2 for analyses), 4-5 = High (numerical value of 3 for analyses), and 6+ = Very High (numerical value of 4 for analyses). Bengtson (2008) reported moderate levels of inter-rater reliability for the RM 2000/V and S (intraclass correlation coefficients of .76 and .72, respectively).

_Psychopathy Checklist-Revised._

The PCL-R (Hare, 2003) is a structured clinical assessment instrument developed to assess psychopathic personality traits. It is comprised of 20 items that are scored on the basis of a semi-structured interview and file review. Evaluators use clinical judgment to rate each item according to the degree to which the item description in the manual matches the personality/behaviour of the individual (0 = No; 1 = Maybe/in some respects; 2 = Yes). Currently, a two-factor structure for the PCL-R is widely recognised and reported in the literature (Factor 1 measures interpersonal/affective features, Factor 2 measures socially deviant lifestyle), although there is evidence to suggest that a three-factor hierarchical model better reflects the multifaceted concept of psychopathy (Cooke, Michie, & Skeem, 2007). In the manual for the most recent revision of the measure, Hare
(2003) describes a two-factor, four-facet hierarchical model. Factor 1 is made up of Facets 1 and 2, measuring interpersonal and affective features, respectively. Factor 2 is made up of Facets 3 and 4, measuring lifestyle and antisocial features, respectively. Psychometric properties are reported in detail in the manual (Hare, 2003). Hare (2003) reported high inter-rater reliability indices for pooled psychiatric samples, weighted by sample size (intraclass correlation coefficients for averaged ratings of .92 to .97).

2.2.3 Procedure

The study was carried out as part of an ongoing program of service evaluation. Ethical approval was obtained from the appropriate committees. Consecutive admissions to the unit from its inception in March 2004 to March 2007 were included. Measures were completed as part of the standard assessment procedures administered to patients during the initial period on an assessment ward. All assessments were conducted with access to file-based information and involved interviews with the patients. Complete background and mental health reports on the patients were available. All assessments were undertaken by multidisciplinary teams of master’s and Ph.D./M.D. level clinicians formally trained in the use of each of the measures and supervised by senior members of the clinical team. The same assessors may have completed more than one of the risk assessment tools or personality measures during the assessment period and in most cases would have had access to the results of all assessments completed to that point. Given that the study was carried out as part of a service evaluation, it was not possible to undertake a second set of independent assessments with the patients in order to examine inter-rater reliability.
Incidents of aggression, including damage to property, verbal aggression, and interpersonal physical aggression directed to staff or other patients were identified by reviewing a register of standard hospital incident forms. The nature and dates of the incidents occurrence were coded by a bachelor’s-level psychology graduate using a modified version of the Overt Aggression Scale (Yudofsky, Silver, Jackson, Endicott, & Williams, 1986). Both a dichotomous variable (incident occurred: yes/no) and a continuous frequency variable (a count of the number of times the incident type occurred during the follow-up) were coded. The period under scrutiny was from the dates of completion of the assessment measures to the end of follow-up period (March 2007) or the date of transfer of the patient to another unit.

A number of procedures were used to control for the differing time periods over which patients were followed. In addition to analyses employing all available data, incidents were recoded using a fixed follow-up period (see Harris, Rice et al., 2003) of 12 months: patients for whom the focal incident (e.g., interpersonal physical aggression) had not occurred and who had been followed up for a total of less than 12 months were removed from the sample, and patients for whom the incident had occurred after 12 months were classified as being incident-free up to that point. This provided an initial, fixed 12-month follow-up period beginning when assessment measures were completed to use in analyses in addition to analyses with all data for the full follow-up period.

4 It was decided to analyse the data using both the full follow-up period and an initial 12-month fixed follow-up period in order to ensure that potentially important findings were not missed. Some studies simply report an average follow-up time and do not attend to the possibility that follow-up times between patients in the sample may vary considerably. Where all patients have been followed-up for a reasonable minimum period this may be an adequate approach. But, if present, differences in follow-up time could confound associations between predictors and outcomes if patients with lower scores on the measures are followed for longer periods than patients with higher scores; incidents of aggression might be missed for higher scoring patients but detected for those with lower scores. A full follow-up period, sometimes referred to in other studies as a ‘variable’ follow-up period, was used in the present study in order to utilize
Another method of controlling for differing follow-up periods involved the use of survival analysis (see below).

2.2.4 Data Analysis

*Associations between assessment measures and outcomes.*

Correlations were calculated to determine associations between assessment measures as well as associations between the assessment measures and outcomes (point biserial correlations, \( r_{pb} \), for dichotomous outcome and Spearman’s rho correlations for frequency count of incidents and for days to first incident). Given the small sample size and associated low statistical power, the point biserial correlations were converted to Cohen’s \( d \) effect sizes using formula 9 in McGrath and Meyer (2006) (the second formula in the appendix of Rice & Harris, 2005) to facilitate interpretation.

*Prediction accuracy.*

Receiver operating characteristics (ROC) analyses were used to investigate the predictive accuracy of the assessment measures for this sample (Mossman, 1994). ROC curves plot the sensitivity (hit rate or true positive probability) of a prediction as a function of specificity (false alarm rate = 1 minus specificity, or false positive probability) (Hanley & McNeil, 1982; Swets et al., 2000). The resulting area under the curve (AUC) is uninfluenced by base rates, and can be interpreted as the likelihood that all available data from the sample of 44 and facilitate comparisons with other studies. An initial 12-month fixed follow-up for an \( n \) of 36 was also used because it controls for unequal follow-up times between individuals. Harris and Rice (2003) have shown that fixed and equal follow-up is one methodological feature that maximizes predictive accuracy. But there is a trade-off if the total sample is small: when the fixed follow-up is short (e.g., a month) most of the sample can be retained but fewer incidents of interest occur. If the fixed follow-up period is longer (e.g., 12 months) more of the focal incidents occur but the sample size can drop (because those followed-up for less than 12 months who are incident-free are removed from the analyses). Associated with this reduced sample size is a loss of statistical power and less confidence in the generalizability of the findings. These concerns likely go some way to explaining differences in findings for individual tools between the two follow-up periods in the present study. Given the need for empirical evaluations of these tools with psychiatric patients in high security settings such as those of the DSPD programme, the use of both follow-ups was considered informative as a first step in this line of research.
the score on a given scale for a randomly selected individual involved in an aggressive incident is higher than that of a randomly selected incident-free individual.

Survival analyses.

Kaplan-Meier survival analysis was used to investigate the relationship between the assessment measures and outcomes over time. The survival function calculated estimates the probability that an individual will survive (i.e., the focal aggressive incident will not occur) for a specified period of time. The survival function can be used to plot a curve of percent survival as a function of time. The resulting survival curve indicates the proportion of the sample (on the y-axis) surviving at any given time (on the x-axis). In research of this kind, the survival curve reflects the cumulative proportion of patients for whom the focal incident has not occurred across the follow-up period.

2.3 Results

The mean number of days from admission to end of the follow-up period was 570.3 (SD = 320.7, Mdn = 611.0, range: 29 to 1095 days). The mean number of incidents of physical aggression resulting in damage to property was .79 (SD = 1.57, Mdn = 0.00, range: 0 to 9). Over 38% of patients damaged property on at least one occasion during the follow-up period. The mean number of incidents of verbal aggression was 10.02 (SD = 12.20, Mdn = 5.00, range: 0 to 58), with 93.2% of the sample exhibiting this type of behavior on at least one occasion. For physical aggression toward another patient or staff the mean number of incidents was .93 (SD = 1.76, Mdn = 0.00, range: 0 to 7), with 38.6% of patients engaging in this behavior on one or more occasions. Due to the fact that almost all of the patients were verbally aggressive during the follow-up period it was
decided to concentrate on the accuracy of the assessment measures in predicting incidents involving damage to property and interpersonal physical aggression only.

2.3.1 Relationships Between the Assessment Tools

Correlations generally indicated moderate correspondence between total scores on the assessment measures with some associations lower than might be expected given that the measures have a number of items tapping the same or similar concepts. Correlations presented in Table 2.1 are for the measures incorporating dynamic variables and their associations with PCL-R Facet scores. The correlations between total scores on the PCL-R, HCR-20, and VRS were all positive and significant. The VRS was also positively and significantly correlated with the RM 2000/V and S scales \( r = .61 \) and \( r = .58, p < .05 \), not tabulated) but the correlations between the RM 2000 scales and the PCL-R and HCR-20 were smaller and nonsignificant \( rs \leq .42, \) not tabulated). Only the Static-99 was significantly correlated with the RM 2000/S \( r = .81, p < .001, \) not tabulated), consistent with the fact that these two tools contain items concerning sexual offending behaviour.

Among the other main findings, the HCR-20 SFRJ was significantly and positively correlated with HCR-20 Total score as well as the Historical and Risk Management scales but the correlation with the Clinical scale was not significant. Correlations between the SFRJ and Factor 1 (and Facet 2) of the PCL-R were also relatively large, positive, and significant. In contrast, the HCR-20 Total score was significantly correlated with all three HCR-20 component scales and PCL-R Total score as well as both Factor 1 (and Facets 1 and 2) and Factor 2. The Clinical scale of the HCR-20 was also positively and significantly correlated with the Dynamic Factors scale and the VRS Total score. The Risk Management scale was also positively and
significantly correlated with the VRS Dynamic Factors scale but unlike the Clinical scale it correlated with Factor 1 the PCL-R as well (and its correlation with Facet 2 approached significance \((r = .33, p = .055)\).

Correlations with VRS Total score were positive and significant for the PCL-R Total, Factor 2 (and Facet 4) scores as well as with HCR-20 Total score and Historical and Clinical scales. The Static Factors scale of the VRS correlated significantly and positively with PCL-R Factor 2 (and Facets 3 and 4), although not with PCL-R Total. In contrast, the Dynamic Factors scale significantly correlated positively with Factor 1 (and Facet 2) and total score on the PCL-R and with all component scales of the HCR-20.
Table 2.1. Correlations Between Component Scores on the HCR-20, VRS, and PCL-R.

<table>
<thead>
<tr>
<th>Instrument/Scale</th>
<th>HCR-20</th>
<th>HCR-20 Total</th>
<th>HCR-20</th>
<th>HCR-20</th>
<th>HCR-20</th>
<th>VRS</th>
<th>VRS</th>
<th>VRS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SFRJ</td>
<td>H-scale</td>
<td>C-scale</td>
<td>R-scale</td>
<td>Total</td>
<td>S-scale</td>
<td>D-scale</td>
<td></td>
</tr>
<tr>
<td>HCR-20 Total</td>
<td>.69***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-scale</td>
<td>.44*</td>
<td>.73***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-scale</td>
<td>.29</td>
<td>.71***</td>
<td>.36*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-scale</td>
<td>.70***</td>
<td>.72***</td>
<td>.28</td>
<td>.35*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VRS Total</td>
<td>.43</td>
<td>.50**</td>
<td>.51**</td>
<td>.40*</td>
<td>.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-scale</td>
<td>-.07</td>
<td>.14</td>
<td>.22</td>
<td>.12</td>
<td>.06</td>
<td>.70***</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D-scale</td>
<td>.71**</td>
<td>.59***</td>
<td>.58***</td>
<td>.39*</td>
<td>.45*</td>
<td>.90***</td>
<td>.35*</td>
<td>-</td>
</tr>
<tr>
<td>PCL-R Total</td>
<td>.18</td>
<td>.54***</td>
<td>.62***</td>
<td>.09</td>
<td>.22</td>
<td>.42*</td>
<td>.28</td>
<td>.42*</td>
</tr>
<tr>
<td>Factor 1</td>
<td>.48*</td>
<td>.50**</td>
<td>.34*</td>
<td>.07</td>
<td>.35*</td>
<td>.19</td>
<td>-.26</td>
<td>.39*</td>
</tr>
<tr>
<td>Factor 2</td>
<td>-.02</td>
<td>.32*</td>
<td>.52***</td>
<td>.21</td>
<td>.09</td>
<td>.46**</td>
<td>.60***</td>
<td>.31</td>
</tr>
<tr>
<td>Facet 1</td>
<td>.30</td>
<td>.36*</td>
<td>.21</td>
<td>-.08</td>
<td>.29</td>
<td>.10</td>
<td>-.24</td>
<td>.24</td>
</tr>
<tr>
<td>Facet 2</td>
<td>.57**</td>
<td>.56***</td>
<td>.43**</td>
<td>.28</td>
<td>.33</td>
<td>.25</td>
<td>-.19</td>
<td>.47**</td>
</tr>
<tr>
<td>Facet 3</td>
<td>-.24</td>
<td>.11</td>
<td>.36*</td>
<td>.07</td>
<td>-.08</td>
<td>.30</td>
<td>.41*</td>
<td>.22</td>
</tr>
<tr>
<td>Facet 4</td>
<td>.06</td>
<td>.28</td>
<td>.38*</td>
<td>.21</td>
<td>.16</td>
<td>.41*</td>
<td>.60***</td>
<td>.22</td>
</tr>
</tbody>
</table>
Note. HCR-20 SFRJ = Structured Final Risk Judgment. HCR-20 H-scale = Historical scale. HCR-20 C-scale = Clinical scale. HCR-20 R-scale = Risk Management scale. VRS S-scale = Static Factors scale. VRS D-scale = Dynamic Factors scale. Spearman rho correlation coefficients are reported between total scores.

*p < .05, two-tailed. **p < .01, two-tailed. ***p < .001, two-tailed.
2.3.2 Predicting Aggressive Behaviour

Reported in the second and third columns of Tables 2.2 and 2.3 is the number within the sample for which each measure was completed as well as mean (and standard deviation) scores. The HCR-20 Historical scale, the VRS Static Factors scale, and the PCL-R were completed for the majority of the sample. Scores on the Clinical and Risk Management scales of the HCR-20 and the Dynamic Factors scale of the VRS were available for approximately three-quarters of the sample. The RM 2000 scales and Static-99 were completed for the 22 participants in the sample with sexual offences.

Two follow-up periods are presented in Tables 2.2 and 2.3, which report AUC values as well as their standard errors (columns 4 and 9) and 95% confidence intervals (columns 5 and 10), Cohen’s $d$ (columns 7 and 12) effect sizes corresponding to the point biserial correlations, and correlations between scores and (i) a dichotomous outcome variable (columns 6 and 11) as well as (ii) a frequency count of aggressive incidents (columns 8 and 13). The dependent variables were incidents involving physical aggression resulting in damage to property (Table 2.2) and interpersonal physical aggression (Table 2.3). The most important findings for the measures are summarized in the following text.
Table 2.2. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the Risk Assessment Tools for Incidents Involving Damage to Property.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>$N$</th>
<th>$\bar{x}$ (SD)</th>
<th>AUC (SE)</th>
<th>95% CI</th>
<th>$\text{Event}^a\ r$</th>
<th>$d^b$</th>
<th>$\text{Freq}^c\ r$</th>
<th>AUC (SE)</th>
<th>95% CI</th>
<th>$\text{Event}^b\ r$</th>
<th>$d^b$</th>
<th>$\text{Freq}^c\ r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCR-20 SFRJ</td>
<td>23</td>
<td>2.4 (0.8)</td>
<td>.73* (.105)</td>
<td>.52-.93</td>
<td>.43*</td>
<td>1.00</td>
<td>.47*</td>
<td>.67 (.126)</td>
<td>.42-.91</td>
<td>.33</td>
<td>.74</td>
<td>.32</td>
</tr>
<tr>
<td>HCR-20 Total</td>
<td>42</td>
<td>30.0 (4.1)</td>
<td>.70** (.085)</td>
<td>.53-.87</td>
<td>.37**</td>
<td>.82</td>
<td>.36*</td>
<td>.60 (.112)</td>
<td>.38-.82</td>
<td>.16</td>
<td>.40</td>
<td>.12</td>
</tr>
<tr>
<td>HCR-20 H-scale</td>
<td>42</td>
<td>16.2 (1.7)</td>
<td>.57 (.095)</td>
<td>.38-.75</td>
<td>.09</td>
<td>.18</td>
<td>.13</td>
<td>.55 (.127)</td>
<td>.30-.79</td>
<td>.05</td>
<td>.12</td>
<td>.02</td>
</tr>
<tr>
<td>HCR-20 C-scale</td>
<td>35</td>
<td>6.1 (1.5)</td>
<td>.66 (.095)</td>
<td>.47-.84</td>
<td>.28†</td>
<td>.59</td>
<td>.24</td>
<td>.71* (.102)</td>
<td>.52-.91</td>
<td>.32*</td>
<td>.76</td>
<td>.30</td>
</tr>
<tr>
<td>HCR-20 R-scale</td>
<td>35</td>
<td>7.2 (2.0)</td>
<td>.77*** (.082)</td>
<td>.61-.93</td>
<td>.44**</td>
<td>1.03</td>
<td>.50**</td>
<td>.66 (.119)</td>
<td>.43-.90</td>
<td>.28</td>
<td>.66</td>
<td>.26</td>
</tr>
<tr>
<td>VRS Total</td>
<td>34</td>
<td>61.2 (7.5)</td>
<td>.72** (.090)</td>
<td>.54-.90</td>
<td>.36*</td>
<td>.79</td>
<td>.39*</td>
<td>.55 (.139)</td>
<td>.28-.83</td>
<td>.08</td>
<td>.19</td>
<td>.05</td>
</tr>
<tr>
<td>VRS S-scale</td>
<td>41</td>
<td>13.2 (3.0)</td>
<td>.67* (.084)</td>
<td>.51-.84</td>
<td>.30*</td>
<td>.63</td>
<td>.38**</td>
<td>.49 (.098)</td>
<td>.30-.68</td>
<td>-.02</td>
<td>-.03</td>
<td>-.02</td>
</tr>
<tr>
<td>VRS D-scale</td>
<td>34</td>
<td>48.0 (6.0)</td>
<td>.68 (.097)</td>
<td>.49-.87</td>
<td>.27†</td>
<td>.58</td>
<td>.30*</td>
<td>.53 (.138)</td>
<td>.26-.80</td>
<td>.06</td>
<td>.13</td>
<td>.02</td>
</tr>
<tr>
<td>RM 2000/V</td>
<td>22</td>
<td>2.9 (0.8)</td>
<td>.74* (.105)</td>
<td>.54-.95</td>
<td>.36†</td>
<td>.91</td>
<td>.42*</td>
<td>.70 (.150)</td>
<td>.41-.100</td>
<td>.23</td>
<td>.76</td>
<td>.24</td>
</tr>
<tr>
<td>RM 2000/S</td>
<td>22</td>
<td>3.1 (0.9)</td>
<td>.72 (.140)</td>
<td>.45-1.00</td>
<td>.34†</td>
<td>.87</td>
<td>.36†</td>
<td>.44 (.248)</td>
<td>.00-.92</td>
<td>-.09</td>
<td>-.29</td>
<td>-.07</td>
</tr>
<tr>
<td>Static-99 (7 bins)</td>
<td>22</td>
<td>4.9 (1.3)</td>
<td>.70 (.133)</td>
<td>.44-.96</td>
<td>.30</td>
<td>.71</td>
<td>.34</td>
<td>.56 (.183)</td>
<td>.20-.92</td>
<td>.07</td>
<td>.15</td>
<td>.10</td>
</tr>
</tbody>
</table>
Note: HCR-20 SFRJ = Structured Final Risk Judgment. HCR-20 H-scale = Historical scale. HCR-20 C-scale = Clinical scale. HCR-20 R-scale = Risk Management scale. VRS S-scale = Static Factors scale. VRS D-scale = Dynamic Factors scale. \textsuperscript{a}Point biserial correlation between scores and dichotomous event occurrence. \textsuperscript{b}Cohen's $d$, converted from the point biserial correlation between scores and dichotomous event occurrence. \textsuperscript{c}Spearman $\rho$ correlation between scores and frequency count of incidents. Data for risk levels rather than total scores are tabulated for the RM 2000 scales and Static-99.

\* $p < .05$, 1-tailed. \** $p < .01$, 1-tailed. \*** $p < .001$, 1-tailed. \textsuperscript{†} $p < .06$
Table 2.3. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the Risk Assessment Tools for Incidents Involving Interpersonal Physical Aggression.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>$n$</th>
<th>$\bar{x}$ (SD)</th>
<th>AUC (SE)</th>
<th>95% CI</th>
<th>Event$^a$ $r$</th>
<th>$d^b$</th>
<th>Freq$^c$ $r$</th>
<th>AUC (SE)</th>
<th>95% CI</th>
<th>Event$^b$ $r$</th>
<th>$d^c$</th>
<th>Freq$^c$ $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCR-20 SFRJ</td>
<td>23</td>
<td>2.4 (0.8)</td>
<td>.80*** (.091)</td>
<td>.62-.98</td>
<td>.54**</td>
<td>1.35</td>
<td>.55**</td>
<td>.75* (.114)</td>
<td>.53-.97</td>
<td>.46*</td>
<td>1.11</td>
<td>.48*</td>
</tr>
<tr>
<td>HCR-20 Total</td>
<td>42</td>
<td>30.0 (4.1)</td>
<td>.68* (.082)</td>
<td>.52-.84</td>
<td>.30*</td>
<td>.65</td>
<td>.26†</td>
<td>.58 (.099)</td>
<td>.39-.78</td>
<td>.14</td>
<td>.31</td>
<td>.10</td>
</tr>
<tr>
<td>HCR-20 H-scale</td>
<td>42</td>
<td>16.2 (1.7)</td>
<td>.48 (.092)</td>
<td>.30-.66</td>
<td>.01</td>
<td>-.02</td>
<td>-.06</td>
<td>.41 (.103)</td>
<td>.21-.61</td>
<td>-.11</td>
<td>-.24</td>
<td>-.15</td>
</tr>
<tr>
<td>HCR-20 C-scale</td>
<td>35</td>
<td>6.1 (1.5)</td>
<td>.68* (.092)</td>
<td>.50-.86</td>
<td>.28*</td>
<td>.62</td>
<td>.21</td>
<td>.58 (.114)</td>
<td>.36-.81</td>
<td>.15</td>
<td>.35</td>
<td>.09</td>
</tr>
<tr>
<td>HCR-20 R-scale</td>
<td>35</td>
<td>7.2 (2.0)</td>
<td>.70* (.090)</td>
<td>.52-.87</td>
<td>.35*</td>
<td>.79</td>
<td>.33*</td>
<td>.73* (.109)</td>
<td>.51-.94</td>
<td>.38*</td>
<td>.91</td>
<td>.34*</td>
</tr>
<tr>
<td>VRS Total</td>
<td>34</td>
<td>61.2 (7.5)</td>
<td>.62 (.096)</td>
<td>.43-.81</td>
<td>.21</td>
<td>.46</td>
<td>.16</td>
<td>.58 (.113)</td>
<td>.35-.80</td>
<td>.19</td>
<td>.41</td>
<td>.10</td>
</tr>
<tr>
<td>VRS S-scale</td>
<td>41</td>
<td>13.2 (3.0)</td>
<td>.50 (.095)</td>
<td>.31-.69</td>
<td>.00</td>
<td>.00</td>
<td>.03</td>
<td>.46 (.102)</td>
<td>.26-.66</td>
<td>-.08</td>
<td>-.18</td>
<td>-.03</td>
</tr>
<tr>
<td>VRS D-scale</td>
<td>34</td>
<td>48.0 (6.0)</td>
<td>.66 (.099)</td>
<td>.47-.86</td>
<td>.27</td>
<td>.59</td>
<td>.21</td>
<td>.63 (.121)</td>
<td>.39-.86</td>
<td>.24</td>
<td>.52</td>
<td>.15</td>
</tr>
<tr>
<td>RM 2000/V</td>
<td>22</td>
<td>2.9 (0.8)</td>
<td>.38 (.148)</td>
<td>.09-.67</td>
<td>-.18</td>
<td>-.43</td>
<td>-.13</td>
<td>.35 (.149)</td>
<td>.06-.64</td>
<td>-.25</td>
<td>-.57</td>
<td>-.23</td>
</tr>
<tr>
<td>RM 2000/S</td>
<td>22</td>
<td>3.1 (0.9)</td>
<td>.57 (.163)</td>
<td>.25-.88</td>
<td>.10</td>
<td>.23</td>
<td>.03</td>
<td>.53 (.172)</td>
<td>.19-.86</td>
<td>.02</td>
<td>.04</td>
<td>-.01</td>
</tr>
<tr>
<td>Static-99 (7 bins)</td>
<td>22</td>
<td>4.9 (1.3)</td>
<td>.33 (.147)</td>
<td>.05-.62</td>
<td>-.22</td>
<td>-.60</td>
<td>-.25</td>
<td>.26 (.153)</td>
<td>.00-.56</td>
<td>-.41†</td>
<td>-1.04</td>
<td>-.40</td>
</tr>
</tbody>
</table>
Note: HCR-20 SFRJ = Structured Final Risk Judgment. HCR-20 H-scale = Historical scale. HCR-20 C-scale = Clinical scale. HCR-20 R-scale = Risk Management scale. VRS S-scale = Static Factors scale. VRS D-scale = Dynamic Factors scale. \( ^a \)Point biserial correlation between scores and dichotomous event occurrence. \( ^b \)Cohen’s \( d \), converted from the point biserial correlation between scores and dichotomous event occurrence. \( ^c \)Spearman \( \rho \) correlation between scores and frequency count of incidents. Data for risk levels rather than total scores are tabulated for the RM 2000 scales and Static-99.

\* \( p < .05 \), 1-tailed. \** \( p < .01 \), 1-tailed. \*** \( p < .001 \), 1-tailed. \( \dagger \) \( p < .06 \)
For both damage to property and interpersonal physical aggression outcomes, the HCR-20 SFRJ was found to have a moderate level of accuracy, with AUCs between .73 to .80 (with large Cohen’s $d$ values corresponding to the point biserial correlations). For damage to property in the initial 12-month follow-up period the effect size was in the medium to large range although the AUC and $r_{pb}$ failed to reach statistical significance. Significant correlations with these dichotomous outcomes ranged from .43 to .54, and significant correlations with the frequency counts ranged from .47 to .55. Using the full follow-up, the HCR-20 Total score also showed a moderate level of accuracy for both outcomes. The Clinical and Risk Management scales similarly predicted both outcomes (although the pattern of significant AUCs according to outcome and follow-up period varied between the two and only the Risk Management scale predicted interpersonal physical aggression for both follow-up periods) with effect sizes generally falling in the medium to large range for all these components. However, the HCR-20 Historical scale failed to predict the outcomes for either follow-up period with small effect sizes at best.

For the outcome damage to property, both VRS Total and Static Factor scale were statistically significant predictors using the full follow-up period (effect sizes in the medium range: AUCs of .72 and .67, respectively; correlations with outcome of .36 and .30, respectively; correlations with frequency counts of .39 and .38, respectively). The AUC and $r_{pb}$ values for the Dynamic Factors scale failed to reach statistical significance although, like the Total and Static Factor scale, its effect size was in the medium range. Neither of the scales nor the Total score predicted interpersonal physical aggression although, for the Total and Dynamic Factor scale again, $d$ values indicated medium effect sizes for this outcome using either follow-up period.
Of the two RM 2000 scales, neither significantly predicted interpersonal physical aggression (AUCs all ns) using either follow-up; negative correlations actually indicating an inverse association with medium effect sizes. However, the RM 2000/V did predict damage to property: AUC = .74) using the full follow-up period, and effect sizes were generally large for both the V and S scales for damage to property.

Both the 4-risk levels and 7-risk levels versions of the Static-99 achieved statistically nonsignificant AUCs for both outcomes using either follow-up. However, for damage to property, Cohen’s $d$ effect sizes for both versions fell in the medium to large range. The same pattern was found for interpersonal physical aggression although for this outcome the correlations were negative, again indicating an inverse association. Static-99 Total score (not included in Tables 2.1 and 2.2) did predict the dichotomous damage to property variable at a statistically significant level using the full follow-up period (AUC = .83, $SE = .13$, 95% CI = .57–1.0; $r_{pb} = .59$, $p < .01$; $d = 1.64$) but, as with its two risk level versions, correlations revealed inverse associations with interpersonal physical aggression using either follow-up.

### 2.3.3 Predicting Repetitive Aggressive Behaviour

An additional series of ROC analyses were carried out to evaluate the accuracy of the measures in predicting 2 or more focal incidents over the full follow-up period. To limit the number of analyses, this series were restricted to the outcome interpersonal physical aggression. Only the AUC for the HCR-20 SFRJ was found to be statistically significant (AUC = .74, $SE = .11$, 95% CI = .52–.95).
2.3.4 Imminence of Outcomes and Failure Rates

One means by which to examine imminence of outcome is to calculate the correlation between assessment measure score and days to first incident of physical aggression (Harris, Rice et al., 2003). Using the initial 12-month follow-up, correlations were generally negative (reflecting fewer days to incident as score increased) and of the same magnitude as the correlations with the dichotomous outcomes. Of note, correlations for the HCR-20 SFRJ ($r = -.48$, $n = 18$) and the HCR-20 Risk Management scale ($r = -.33$, $n = 28$), were significant for the outcome interpersonal physical aggression ($p < .05$, 1-tailed).

A series of Kaplan-Meier survival analyses were carried out to investigate whether higher scores on the measures were associated with faster time to an incident involving aggression. Again, to limit the number of analyses, this series was restricted to the outcome interpersonal physical aggression. For these analyses, the focus was on the three scales composed of dynamic variables: the Clinical and Risk Management scales of the HCR-20, and the Dynamic Factors scale of the VRS. Although the HCR-20 SFRJ score predicted interpersonal physical aggression as well, it was not included in these analyses because, of the 23 patients with a SFRJ, all eight of the individuals involved in incidents of interpersonal physical aggression had a rating of 3.

Scores on each of the measures were divided into four groups based on quartiles and the survival functions for groups plotted. For the HCR-20 Risk Management scale, the difference between failure rates for the first (scores < 6) and fourth quartile (scores $\geq 9$) was significant (Log Rank test: $\chi^2 = 4.012$, $p = .045$), with the fourth quartile group showing the fastest failure rate compared with the other groups (see Figure 2.1). No
other comparisons revealed significant differences. At the 12-month mark, the cumulative proportion surviving in the first, second, third, and fourth quartile groups were 100%, 75%, 83%, and 47%, respectively.

The same procedure was followed for the HCR-20 Clinical scale and the VRS Dynamic Factor scale. For the Clinical scale, at the 12-month mark no linear association was evident; the cumulative proportion surviving in the first, second, third, and fourth quartile groups were 81%, 67%, 83%, and 67%, respectively. Somewhat clearer was the pattern for the VRS Dynamic Factors scale: 56% and 58% of the third and fourth quartile groups were still incident-free compared with 86% and 75% of the first and second quartile groups, suggesting that a division at the median value provided the clearest distinction in failure rates. Overall, however, for neither the Clinical scale nor the Dynamic Factors scale were the differences in failure rates for quartile-based groups statistically significant, nor were significant differences in failure rates found when two groups were used, divided at the median score for each measure.
Figure 2.1. Interpersonal physical aggression survival curves for HCR-20 Risk Management (RM) quartile groups.
2.4 Discussion

Consistent with the intake criteria for the DSPD service, one of the most striking findings of the study is that patients’ scores on the risk assessment tools were notably higher than those reported elsewhere in the literature. Focusing first on the HCR-20, the mean Total score for the present sample was 32% higher than mean HCR-20 Total scores reported by Macpherson and Kevan (2004) for their sample of 86 male maximum security psychiatric inpatients in Scotland, 32% higher that that reported by Dolan and Fullam (2007) for their sample of 136 male medium security psychiatric inpatients in England, and 30% higher than that reported by de Vogel and de Ruiter (2005) for their sample of 42 male forensic psychiatric patients in The Netherlands.

The same pattern of higher scores was evident for the other assessment measures also. The mean VRS Total score in the present sample was 33% higher than that reported by Dolan and Fullam (2007), 43% higher than that reported by Grevatt et al. (2004) for their sample of 44 male psychiatric inpatients detained in an independent security facility in England, and 32% higher than that reported by Wong and Gordon (2006) for their pooled sample of 918 male offenders drawn from the Canadian correctional federal and provincial systems and a forensic psychiatric facility.

Research on the accuracy of the Static-99 and Risk Matrix 2000 scales at predicting aggression among forensic psychiatric inpatient samples has not yet been reported in the peer-reviewed literature but, looking at available community follow-up studies with forensic psychiatric samples, it is clear than the mean scores for these tools in the present sample are numerically higher than reported elsewhere. For example, the highest mean total score on the Static-99 reported in the peer-reviewed literature is 4.95
(SD = 2.1) for a sample of 258 male sexual offenders assessed while incarcerated in a maximum security psychiatric hospital in Canada (Looman, 2006), a score 15% lower than that of the present sample (mean total score = 5.82, SD = 2.5). For the Risk Matrix 2000/V and S scales, Bengtson (2008) reported mean risk level scores of 2.57 and 2.39, respectively, for a Danish sample of 304 male sex offenders who underwent a forensic psychiatric evaluation, 10% and 22% lower than the respective mean risk level scores in the present sample.

2.4.1 Prediction Accuracy

An important objective for the present study was to determine if the risk assessment tools completed in the course of normal clinical practice would have a reasonable level of predictive validity in a forensic inpatient psychiatric sample for which scores were expected to have a restricted range that was skewed high. Although the prospective design is a strength of the study given this objective, the small sample size, large number of analyses conducted, and lack of inter-rater reliability for the measures are all limitations that mean the findings must be viewed as preliminary. Starting with the HCR-20, findings for the outcome damage to property were generally consistent with those reported by Gray et al. (2003), with medium to large effect sizes although AUC values were numerically lower in the current study. The Total score, Risk Management scale, and SFRJ all predicted this outcome (AUC values ≥ .70) over the full follow-up and the Clinical scale predicted it using the initial 12-month follow-up.

For prediction of interpersonal physical aggression, results for the HCR-20 were also generally consistent with those reported elsewhere. With the exception of the Historical scale, all components of the HCR-20 were significant predictors in the study
sample (AUCs $\geq .68$, with small to large effect sizes). Of particular note, patients with the highest quartile of scores on the Risk Management scale showed the fastest failure rate among the sample. Survival analyses, however, using the Clinical scale and the VRS Dynamic scale failed to reveal significant differences in failure rates. Few other studies have employed survival analytic techniques with HCR-20 components despite its value as a means to investigate imminence of an event.

Belfrage, Fransson, and Strand (2000) reported similar findings to those presented here for a sample of 41 male offenders (30 of whom were diagnosed as psychopaths using the Psychopathy Checklist: Screening Version (PCL:SV: Hart, Cox, & Hare, 1995) incarcerated in two maximum-security correctional facilities in Sweden. Gray et al. (2003) reported significant indices for a Historical plus Clinical scales Total (Risk Management items were not scored and SFRJs were not reported), as well as for the Historical, and Clinical scales. As an aside, it is worth noting here that for physical aggression Gray et al. found the Brief Psychiatric Rating Scale (BPRS: Overall & Gorham, 1962) had a numerical larger AUC (.84) than did the Clinical scale (.79); the AUC of .68 for the Clinical scale for interpersonal physical aggression in the present study compares even less favourably, suggesting that assessment of clinical concerns in DSPD patients for risk assessment purposes may be improved upon.

Dolan and Fullam (2007) reported that all HCR-20 components were significantly predictive (although they too did not include SFRJs). Grevatt et al. (2004), who also did not include data for SFRJs, reported significant AUC values for the Clinical scale in predicting an inclusive outcome of violence, abuse or harassment as well as 3 or more incidents involving physical assault. In the present sample, only the AUC value of .74
for the SFRJ among the risk assessment tools was statistically significant when the outcome was repetitive interpersonal physical aggression (2 or more incidents during the follow-up).

HCR-20 SFRJs have already been shown to predict violence among forensic psychiatric patients in a community setting (Douglas et al., 2003), and the present data support this approach with forensic psychiatric inpatients too. Similarly, others have reported that structured risk judgments made by multidisciplinary teams in forensic inpatient settings have predictive validity, at least over relatively short periods (Fuller & Cowan, 1999; more generally see McNeil & Binder, 1991). However, further investigation of the risk assessment process carried out in this manner is required.

Determining what factors or characteristics of the patient and circumstances are given most weight by the assessment team should enable improvements in prediction accuracy (through refinement of the tools structuring the approach) and in training (through identifying and then countering clinician biases). There is also the issue of whether such risk assessments can have a self-fulfilling effect. In community settings, this effect may result through increased detection rates for violence among ‘flagged’ patients. In a high security setting with constant surveillance, rather than improved detection, the effect might result from the increased attention and efforts by staff, which should (ideally) be intended to manage risk but are quite possibly experienced negatively by patients and met with increased rates of aggressive behaviour.

Given that scores on the Clinical and Risk Management scales have been shown to change during periods of institutionalization (Belfrage & Douglas, 2002), evaluating the utility of multiple administrations of these scales and repeatedly revised SFRJs with
the DSPD sample for detecting lowered risk over time (and as part of a treatment
evaluation strategy, see Langton, 2007) is an important next step (cf. Belfrage, Fransson,
& Strand, 2004). Indeed, this represents one of the key challenges for the field more
generally, with dynamic risk factors central to assessment of intra-individual risk over
time (Douglas & Skeem, 2005) and quite distinct in practical terms from the
determination of risk status that can be effectively undertaken on the basis of static risk
factors (Heilbrun, 1997).

Of course, the Dynamic Factors scale of the VRS is another promising tool in this
regard. However, the picture that emerges regarding the VRS is not consistent across
outcomes or studies. For the present sample, the VRS Total and Static Factor scale
scores were significantly correlated with damage to property over the full study period
with the correlation for the Dynamic Factor scale approaching significance; $r_{pb} \geq .27$,
representing a medium effect size. The AUC value for the Dynamic Factor scale just fell
short of statistical significance but the Total and Static Factor scores were significant.
Importantly, no component of the VRS significantly predicted interpersonal physical
aggression for either the full or 12-month follow-up (although medium effect sizes were
again found for the Total and Dynamic Factor scale), and failure rates for high- versus
low-risk groups revealed no significant differences contrary to findings concerning post-
release recidivism reported by Wong and Gordon (2006). Grevatt et al. (2004) found that
the VRS predicted neither damage to property nor interpersonal physical aggression in
their sample over a 6-month period. In contrast, in a study looking only at physical
assaults on others, Dolan and Fullam (2007) reported significant results for the Total and
Dynamic Factors scale scores over a 12-month period. As Wong and Gordon (2006)
note, evidence that scores on the Dynamic Factors scale change over time and that such changes are associated with changes in risk for violent or other criminal behaviour has yet to be reported in the peer-reviewed literature (although the developers’ unpublished data are encouraging; see Wong, Olver, & Stockdale, 2009, for discussion of their work). Further evaluation in a variety of jurisdictions with patients/offenders is clearly required given the intended use of the VRS as a ‘treatment friendly’ tool with violence-prone individuals with personality disorders (Wong, Gordon, & Gu, 2007).

Of the various risk assessment tools, risk levels for the Static-99 and RM 2000/S were not statistically significant predictors of interpersonal physical aggression. However, the AUC value for the RM 2000/V was statistically better than chance at predicting damage to property over the full follow-up; it also had statistically significant medium size correlations with this outcome and with the frequency count of such incidents, and large effect sizes were found for both the V and S scales using the full follow-up period. Similarly, a large effect size was found for the Static-99 Total score for this outcome. It seems likely that one important reason for the mixed performance of these three scales is the fact that none of them were designed to inform the assessment of risk for violence in psychiatric inpatient settings (Hanson & Thornton, 1999; Thornton, 2005). Instead, the correctional/prison samples used to develop and test these measures were followed up in community settings for a period of years after release (Hanson & Thornton, 2000; Thornton et al., 2003).

Furthermore, although the Static-99 and RM 2000/V were intended to provide an assessment of risk for violence, an outcome that encompasses the interpersonal physical aggression considered in the present study, the RM 2000/S was designed to predict sexual
recidivism, an outcome for which there was no equivalent in the present study. As well, the subset of patients with sexual offences in the present sample \( n = 22 \) and for whom the Static-99 and RM 2000 scales could therefore be completed was very small, limiting generalisability and affording low statistical power for evaluation purposes by conventional standards. As such, these specific results must be treated with particular caution. It remains to be determined if scores on these risk assessment tools accurately predict adverse events in other forensic psychiatric inpatient settings. Certainly, their exclusive reliance on static information renders them without clear utility for informing patients’ treatment plans or short-term risk management strategies.

Despite the accuracy data for the HCR-20 for both outcomes and for the VRS for damage to property appearing generally consistent with previous studies of forensic psychiatric inpatients, the findings could still be viewed as somewhat underwhelming. Clearly, ranking DSPD patients in terms of their risk of engaging in interpersonal physical aggression or causing damage to property can be undertaken with these measures performing statistically significantly better than chance and yielding small to large effect sizes. But other risk assessment procedures have numerically higher AUC values yet have rates of false positives at various cut-levels or risk bands (Banks et al., 2004; Monahan et al., 2005) that underscore the challenges inherent in clinical decision making with individual patients. The questions addressed by the present study do not speak to the issue of whether individuals entering the DSPD programme are being detained ‘unnecessarily’ (it is hard to see, given the current state of the field, what alternative procedures for identification might be adopted for this policy initiative). But the findings do indicate that some of the tools may have no practical utility and limited
predictive validity once DSPD offenders/patients have been removed from the larger UK offender-plus-patient populations and placed in the DSPD units. Further evaluation of these measures in the DSPD setting are called for but so too are investigations of alternative tools and methods of appraising risk and change in risk over time. Functional assessments have proven highly useful for understanding the causes or precipitants of aggression in individual children, for informing intervention efforts to reduce their violent behaviours, and for evaluating the effects of these intervention efforts (Fox & Gable, 2004). The promise of such procedures with adult forensic populations has yet to be realized (Daffern, Howells, & Ogloff, 2006), with the DSPD programme a seemingly ideal place to start.

2.4.2 Relationships Between Measures

Relatively few studies have examined the concurrent validity of the Static-99 and RM 2000 scales. Correlational analysis using the total scores of the other measures revealed few significant associations with either (data not tabulated). The only significant correlation for the RM 2000/V was with the VRS Total. The only significant correlations for the RM 2000/S were with the VRS Total and Static-99 (an association with the latter would be expected given that both were designed to predict sexual recidivism specifically and have similar item content). As already mentioned, however, the small number of sexual offenders in the present sample renders such results preliminary only, with some of the larger correlations likely failing to reach significance because of the low power afforded by the small sample size.

In contrast, there were a range of significant associations among the other measures although it should be kept in mind that an unknown portion of the concurrent
validity coefficients could be due to common assessor variance and biases because assessors completed more than one of the measures in some cases and had access to those completed by others during the assessment process. The correlation between total scores on the PCL-R and the HCR-20 were generally positive and significant although of a smaller magnitude to those reported elsewhere (e.g., Douglas & Webster, 1999; Gray et al., 2003). Of note, however, in the present study the HCR-20 Clinical scale was not significantly correlated with PCL-R Total, Factor, or Facet scores, suggesting perhaps that some of the clinical concerns of this sample are distinct from personality features. It is interesting that the highest correlation for the HCR-20 SFRJ with PCL-R components was with Factor 1 (and Facet 2), suggesting that assessors were most heavily influenced by the affective features of psychopathy when making their risk appraisal. The HCR-20 Risk Management scale was significantly associated with PCL-R Factor 1, suggesting that the interpersonal and affective features were similarly influential in rating these items. It is perhaps not surprising that a structured risk assessment based in part on interviews with the patients would reflect these features given their likely salience. It could be seen as fortunate then that in the DSPD setting it is these arrestive features that have predictive validity also (see chapter 3).

The general pattern of correlations between scores on the VRS and PCL-R was similar to that reported by Wong and Gordon (2006) although the size of associations was again smaller in the present study. The notable exception and anomaly being that the VRS Static Factor score was actually negatively, though not significantly, correlated with the PCL-R Factor 1 in the present sample. The Dynamic Factors scale was significantly correlated with Factor 1 (and Facet 2) of the PCL-R, perhaps again reflecting the
influence of affective features of psychopathy when assessing criminogenic needs. Compared with the correlations reported by Dolan and Fullam (2007) between the VRS and HCR-20, associations between these measures in the present study were generally smaller but still positive and significant. Again, the exception was the VRS Static Factors scale, which was not significantly associated with any HCR-20 component score in the present study. The Dynamic Factors scale was significantly correlated with all components of the HCR-20. Its large correlation with the HCR-20 SFRJ indicates that criminogenic needs were also weighed heavily in this global or summary risk appraisal relative to static or historical features. Correlations between the Dynamic Factors scale and the HCR-20 Clinical and Risk Management scales indicate some overlap in their coverage of dynamic predictors providing the basis for treatment planning.

Overall, the patterns of association provide modest support for claims of the concurrent validity of the measures (with the strongest relationships appearing for the HCR-20, VRS, and PCL-R), generally consistent with findings reported elsewhere. The small sample size demands caution in interpreting differences in magnitude. With that in mind, it is still of interest to note that in this DSPD (i.e., high scoring) sample, the lower magnitudes indicate that each measure possesses a fair amount of independent variance, which is promising in terms of the possibility that some might have incremental predictive validity within a comprehensive risk assessment and management approach. However, their incremental validity in statistical prediction models remains to be tested with larger samples. This issue and that of evaluating the association between changes in dynamic risk scores and changes in institutional infractions over time represent the next phases of the current project at the Peaks Unit. Ultimately, it is hoped that reductions in
risk in DSPD patients will be empirically linked with treatment efforts and so provide the basis for decisions concerning their transfer and eventual release.
CHAPTER 3: Study #2

Personality traits as predictors of inpatient aggression in a high security forensic psychiatric setting: Prospective evaluation of the PCL-R and IPDE dimension ratings.

Abstract

The Dangerous and Severe Personality Disorder (DSPD) initiative in England and Wales provides specialized care to high risk offenders with mental disorders. This study investigated the predictive utility of personality traits, assessed using the Psychopathy Checklist-Revised and the International Personality Disorder Examination, with 44 admissions to the DSPD unit at a high security forensic psychiatric hospital. Incidents of interpersonal physical aggression were observed for 39% of the sample over an average 1.5 year period following admission, and a similar number of incidents occurred involving damage to property. Only Borderline PD dimension scores predicted damage to property. Histrionic PD predicted interpersonal physical aggression, and Histrionic, Borderline, and Antisocial PDs all predicted repetitive (2+ incidents of) interpersonal physical aggression. PCL-R Factor 1 and Facets 1 and 2 were also significant predictors of interpersonal physical aggression. PCL-R Factor 1 and Histrionic PD scores were significantly associated with imminence of interpersonal physical aggression. Results were discussed in terms of the functional link criterion for DSPD service entry and the utility of personality traits in risk assessment and treatment of specially selected high risk forensic psychiatric patients in secure settings.

3.1 Introduction

The Dangerous and Severe Personality Disorder (DSPD) initiative in England and Wales, underway since 2000 (Home Office and Department of Health, 1999), is intended to provide specialized care to high risk offenders with mental disorders (Maden, 2007). The DSPD service is a collaboration between the Department of Health, the Ministry of Justice (formerly Home Office) and HM Prison Service, involving approximately 300 places in prisons and high security hospitals at four sites. The purpose of these services (officially recognized as “pilots”) is to manage and treat individuals meeting three criteria: (i) a finding of a severe disorder of personality, (ii) a high risk of serious harm to others (more likely than not to commit an offence that might be expected to lead to serious physical or psychological harm from which the victim would find it difficult or impossible to recover), and (iii) a functional link between the disorder and the risk of re-offending (DSPD Programme et al., 2008, p. 8 and pp. 14-15).

The first criteria is problematic given the marked lack of consensus regarding what constitutes severe personality disorder (Blackburn, 2000; Tyrer, 2004), and there is a lack of specificity regarding the second and third criteria also. Nevertheless, operational definitions have been provided and the use of risk assessment tools and measures of personality traits are clearly indicated. A number of these tools have been included in the prescribed DSPD assessment battery in order to inform decisions about entry to the service and treatment needs and to facilitate the measurement of change (DSPD Programme et al., 2008; Duggan & Howard, 2009). The battery includes the Historical Clinical Risk - 20 (HCR-20: Webster et al., 1997), Violence Risk Scale (VRS: Wong & Gordon, 2000, 2006), the Static-99 (Hanson & Thornton, 2000; Harris, Phenix
et al., 2003), the Risk Matrix 2000 Scales (RM 2000/Sexual and RM 2000/Violence scales (Thornton, 2005; Thornton et al., 2003), the Psychopathy Checklist-Revised (PCL-R: Hare, 2003), and the International Personality Disorder Examination (IPDE: Loranger et al., 1994). The first four of these are risk assessment instruments (the criterion predictive validity of which was evaluated in the previous chapter) and the latter two are measures of personality traits.

Although the selection of measures in the DSPD battery reflects the growing body of psychometric data reported in the peer-reviewed literature for most of them, there remain significant gaps in what is known about their utility for distinct populations, in particular settings, and over specified timeframes. The two measures of personality traits (the PCL-R and the IPDE), are the subject of the present study. Given their selection for use within the DSPD service, one important question, reflecting a concern with staff and patient safety and institutional security (and germane to the second criterion concerning risk of future harm), is whether or not these measures predict types of institutional infractions, particularly physically aggressive behaviour, for personality disordered patients in a high security forensic psychiatric setting. It is with this question of context-specific practical utility that the current study is principally concerned.

3.1.1 Pertinent Findings for the Psychopathy Checklist

The PCL-R (Hare, 2003) and its derivatives are relatively well established predictors of institutional and post-release violence and criminal recidivism among correctional and forensic psychiatric populations across cultures (see, for example, Coid et al., 2007; Leistico et al., 2008). However, the small effect sizes, particularly for institutional misconduct (Campbell et al., 2009; Walters, 2003), and the statistically
significant variation in effect sizes across studies within institutional settings (Guy, Edens, Anthony, & Douglas, 2005) suggests that predictive validity for the DSPD population cannot be assumed. Indeed, despite the broader evidence base generally in support of the PCL-R, the intake criteria specified for the DSPD service would be expected to produce an attenuated range of scores with the majority of patients falling in the moderate to high range. This could quite possibly impact the PCL-R’s predictive validity within DSPD settings. Given the role afforded PCL-R assessments in policy and legislative initiatives outside the UK also (for example, the Dangerous Offender Law in Canada [Eaves, Douglas, Webster, Ogloff, & Hart, 2000; Trevethan, Crutcher, & Moore, 2002], and various civil commitment statutes across the US [Doren, 2002; Jackson & Richards, 2007], as well as the influence of PCL-R assessments on management and treatment planning including denial of treatment [Edens & Petrila, 2006]), this concern is likely shared by professionals using the PCL-R in similar high security forensic mental health and criminal justice settings with specially selected populations (i.e., personality disordered, high risk psychiatric patients/offenders).

As well as this question concerning its applied utility, at least in terms of risk assessment within high security settings, of additional interest is the differential predictive validity of the PCL-R’s component factors. A two-factor structure for the PCL-R is widely recognised and reported in the literature. The first factor reflects the interpersonal/affective features of psychopathy while the second factor reflects the socially deviant lifestyle aspects (Hare, 1996). There has been, however, considerable debate about the structure of psychopathy as operationalized by the PCL-R (Cooke et al., 2007; Neumann, Hare, & Newman, 2007; Patrick, Hicks, Nichol, & Krueger, 2007) with
three- and four-factor models being advanced. In the manual for the most recent revision of the measure, Hare (2003) describes a two-factor, four-facet hierarchical model. Factor 1 is made up of Facets 1 and 2, tapping interpersonal and affective features, respectively. Factor 2 is made up of Facets 3 and 4, measuring lifestyle and antisocial features, respectively. The first three of these four facets are the same as the three factors in Cooke and Michie’s (2001) model (which Cooke & Michie labelled arrogant and deceitful interpersonal style, deficient affective experience, and impulsive and irresponsible behavioural style). Considerable data has accrued regarding the predictive validity of PCL-R Total and Factor 1 and 2 scores.

Findings from meta-analyses indicate that Factor 2 effect sizes are larger than Factor 1 effect sizes (e.g., Gendreau, Goggin, & Smith, 2002; Guy et al., 2005; Hemphill, Hare, & Wong, 1998; Leistico et al., 2008; Walters, 2003) but the magnitude of the difference has varied across studies and, importantly for the present study, according to type of institutional infraction. Drawing on 95 nonoverlapping studies, Leistico et al. (2008) reported mean weighted effect sizes (Hedges’ d) of 0.53 for PCL Total (n = 6,137), 0.41 for Factor 1 (n = 3,898) and 0.51 for Factor 2 (n = 3,848) for an inclusive category of institutional infractions. Guy et al. (2005) looked at 38 independent samples and reported mean weighted effect sizes ($r_w$) of .26 for PCL Total (n = 2,477), .20 for Factor 1 (n = 1,073) and .24 for Factor 2 (n = 1,073) for a combined verbal/destruction outcome. For physical violence, the mean weighted effect sizes were .17 for PCL Total (n = 3,502), .14 for Factor 1 (n = 2,129) and .15 for Factor 2 (n = 2,129).

Among relatively recent individual studies of forensic psychiatric inpatients the findings have varied. For a sample of 34 mentally disordered offenders in a UK medium
security facility, Gray et al. (2003) found that PCL-R Total and Factor 2 scores were significantly correlated with violence to property ($r = .38, p < .05$ and $.58, p < .001$, respectively) and physical aggression ($r = .35, p < .05$ and $.36, p < .05$, respectively), but Factor 1 scores were not for either outcome ($r \leq .20, ns$). Similarly, among 42 personality disordered males detained in a Dutch forensic psychiatric setting, de Vogel and de Ruiter (2005) reported that Total and Factor 2 scores were significantly correlated with inpatient violence ($r = .42, p < .01$ and $.58, p < .01$, respectively) but, again, Factor 1 scores were not ($r = .24, ns$). Certainly, the lower statistical power afforded by the size of these samples is an issue but the difference in effect sizes for Factor 1 compared with Factor 2 and PCL-R Total is noteworthy nevertheless. More consistent with the effect sizes reported in the Guy et al. (2005) meta-analysis, McDermott, Edens, Quanbeck, Busse, and Scott (2008) found significant correlations between aggression towards staff and Total and Factor 2 scores ($r = .18, p < .05$ and $.21, p < .05$, respectively) but a nonsignificant correlation for Factor 1 ($r = .16, ns$) for their US sample of 108 forensic psychiatric patients. In that study, Factor 2 also predicted aggression against other patients ($r = .18, p < .05$), although interestingly correlations with a combined aggression category were not significant for Total, Factor 1 or Factor 2 scores ($rs = .09, .08$, and $.12, ns$).

In addition to mixed results in the broader literature concerning the superior predictive validity of the Factor 2 compared with Factor 1, meta-analytic results have shown the country where studies were carried out to be a significant moderator. Consistent with the results of the individual studies mentioned above, both the Guy et al. (2005) and Leistico et al. (2008) meta-analyses reported larger effect sizes for PCL Total
and Factor 1 and 2 scores for studies conducted in European countries with effect sizes in Canadian studies generally somewhat smaller and those in US studies smaller still. Guy et al. identified quality of PCL assessments in US contexts, differences in the racial and ethnic composition of samples between countries, and possible fundamental differences in the management of institutions between countries as possible explanations worthy of further investigation. Regardless, this main finding suggests that larger effect sizes might be expected in studies carried out in non-US institutions such as the setting of the present investigation.

In contrast to the large number of studies reporting on Factor 1 and 2, only a few studies have examined the predictive validity of the PCL-R facets. Walters, Knight, Grann, and Dahle (2008) conducted a meta-analysis using six studies (encompassing samples of civil psychiatric patients and prison inmates) and found that all facet scores achieved a significant mean effect size for both general and violent post-release recidivism (outcomes encompassing any charge or conviction and violent acts, charges or convictions), although the effect size for Facet 4 (antisociality) was significantly higher than that for Facets 1 and 2 for general recidivism and those for Facets 1, 2, and 3 for violent recidivism. Based on these data it seems reasonable to hypothesize that all four facets would predict incidents of aggression among DSPD inpatients also.

3.1.2 Pertinent Findings for the International Personality Disorder Examination

The IPDE (Loranger et al., 1994) is a semi-structured clinical interview developed to assess personality disorders in a manner compatible with the International Classification of Diseases (ICD) and Diagnostic and Statistical Manual (DSM) classification systems. Unlike the PCL-R, the predictive validity of the IPDE has yet to
receive concerted empirical attention. However, based on factor-analytic studies using
the IPDE one can tentatively postulate that certain dimension scores would predict
incidents of aggression in secure settings. Ullrich and Marneros (2004) administered the
IPDE to 105 offenders and 80 non-criminal controls and found the same three-factor
structure in both samples. One factor had high negative loadings on dependant and
anxious personality disorders. Another factor had a high negative loading for schizoid
personality disorder and was defined by anankastic (comparable to obsessive compulsive)
features. More importantly for the present study, the first factor was comprised of
emotionally unstable, histrionic, paranoid, and dissocial/antisocial traits and, for
offenders, high scores on this factor were strongly associated with number of prior
violent offences as well as with self-reported hostility (see also Ullrich & Marneros,
2007).

Howard, Huband, Duggan, and Mannion (2008) reported similar results using
IPDE scores with a community sample of 224 treatment-seeking patients. Second-order
factor analysis revealed a three-factor structure, with Histrionic-Narcissistic, Antisocial,
and Conduct Disorder primary factors characterizing their second-order Antisocial factor
along with smaller loadings on the Borderline and Paranoid factors. Of the three second-
order factors, it was this Antisocial factor that had a distinct set of positive associations
with self-report measures of anger expression, impulsivity, alcohol and substance abuse,
history of aggression and criminal as well as specifically violent convictions and
custodial sentences.

Howard et al.’s (2008) results were also broadly consistent with those of an earlier
study with 167 forensic psychiatric patients, in which Blackburn and Coid (1998)
described a four-factor model using DSM-III PD dimensionalized scores (made with the Structured Clinical Interview for DSM-III Axis II disorders [SCID-II: Spitzer & Williams, 1983]). Histrionic, narcissistic, antisocial, borderline, and passive-aggressive personality traits all loaded on a factor that strongly correlated with PCL-R Total and Factor 1 and 2 scores (which, as reviewed above, are established predictors of violent behaviour), as well as with various criminal history variables. Blackburn, Logan, Renwick, and Donnelly (2005) reported a series of exploratory and confirmatory factor analyses of IPDE data from 168 forensic psychiatric patients. They identified two superordinate factors, dividing the PDs into clusters, which the authors’ labelled “Anxious-Inhibited” and “Acting Out.” The latter included antisocial and histrionic PDs and was significantly correlated with the PCL-R.

A key feature of personality disorders is difficulty with interpersonal relationships and high levels of conflict (American Psychiatric Association, 2000; Johnson, Rabkin, Williams, Remien, & Gorman, 2000) and an association between personality disorders and violent behaviour in adults as well as adolescents has been demonstrated although little of this work has been completed with forensic patient or offender samples (Coid et al., 2006a,b; Johnson, Cohen, et al., 2000; McMurran & Howard, 2009; Miller, Zadolinnyi, & Haffner, 1993; Moran et al., 2003; Tardiff, Marzuk, Leon, & Portera, 1997; Widiger & Trull, 1994; Wormith, Olver, Stevenson, & Girard, 2007). Based on the factor analytic research with forensic psychiatric patients and offender samples reviewed above, as well as these findings concerning personality disorders more generally, it seems reasonable to hypothesize that the IPDE dimension scores for antisocial, borderline, histrionic, and narcissistic PDs would predict inpatient aggression.
Writing specifically about the PCL scales (but of equal relevance also to other measures considered relevant to predicting violence), Edens (2006) cogently argued that the research to date should “compel examiners to be very specific about exactly what types of settings and circumstances are being considered when making statements about risk for future violence” (p. 61). Clearly, demonstration of predictive validity for the PCL-R and the IPDE dimension scores over specified time periods with individuals who have personality disorders and are deemed to be at high risk of aggression would be of considerable interest and practical utility to professionals working in the DSPD programme and other criminal justice and mental health settings. It was hypothesized that PCL-R Total, Factor, and Facet scores would predict various types of institutional aggression as would the four IPDE Cluster B personality disorder dimension scores. Given the relatively limited research on the predictive validity of PCL-R Facet and IPDE dimension scores, no a priori hypotheses concerning relative superiority were advanced.

3.2 Method

3.2.1 Setting and Participants

See pages 18 and 19 for a description of the Peaks Unit and the 44 admissions comprising the sample.

3.2.2 Measures

The Psychopathy Checklist-Revised (PCL-R: Hare, 2003) is a structured clinical assessment instrument developed to assess psychopathic personality traits. It is comprised of 20 items that are scored on the basis of a semi-structured interview and file review. Evaluators use clinical judgment to rate each item according to the degree to which the item description in the manual matches the personality/behaviour of the
individual (0 = No; 1 = Maybe/in some respects; 2 = Yes). Psychometric properties are reported in detail in the manual. Hare reported high inter-rater reliability indices for pooled samples, weighted by sample size (intraclass correlation coefficients for averaged ratings of .92 to .97).

The International Personality Disorder Examination (DSM-IV module) (IPDE: Loranger et al., 1994) is a semi-structured clinical interview developed to assess personality disorders in a manner compatible with the International Classification of Diseases (ICD) and Diagnostic and Statistical Manual (DSM) classification systems. Scoring of the 93 items is undertaken using a 3-point rating (0 = absent or within normal range; 1 = present to an accentuated degree; 2 = pathological/meets criterion) and the final algorithmic integration of the items permits both diagnostic classifications and dimensional scores for Antisocial PD, Borderline PD, Histrionic PD, and Narcissistic PD, (as well as Paranoid PD, Schizoid PD, Schizotypal PD, Avoidant PD, Dependent PD, and Obsessive-Compulsive PD although these latter PDs were not the subject of the present investigation). Satisfactory psychometric properties have been reported by Loranger and colleagues (Loranger et al., 1994; Loranger, Janca, & Sartorius, 1997). Indices of inter-rater reliability reported elsewhere have been acceptable (intraclass correlation coefficient = .90, Ullrich & Marneros, 2004).

3.2.3 Procedure

See pages 24 to 26 for a description of the procedure.
3.2.4 Data Analysis

See pages 26 to 27 for a description of the data analytic strategies used.

3.3 Results

For mean follow-up time and percentage of the sample committing types of aggressive acts, see pages 27 to 28.

3.3.1 Relationships Between the Assessment Measures

Inter-measure correlations are presented in Table 3.1. Associations between components of the PCL-R Total score and the two Factor scores were medium to large and statistically significant ($r_s \geq .58$, $p < .001$), as were correlations between each Factor and its constituent Facets ($r_s \geq .77$, $p < .001$). Factor 1 and Factor 2 were not significantly correlated with each other. Associations between discrete IPDE personality disorder Cluster B dimension scores were significant and generally medium in size ($r_s = .38-.50$). Inter-measure correlations revealed only a few significant associations: the Antisocial PD dimension score was correlated with Factor 2 ($r = .46$) and Facet 4 of the PCL-R ($r = .49$).
Table 3.1. Correlations Between Measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>PCL-R Total, Factors and Facets</th>
<th>IPDE Cluster B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>PCL-R Factor 1</td>
<td>.67***</td>
<td>-</td>
</tr>
<tr>
<td>PCL-R Factor 2</td>
<td>.58***</td>
<td>-.10</td>
</tr>
<tr>
<td>PCL-R Facet 1</td>
<td>.55***</td>
<td>.92***</td>
</tr>
<tr>
<td>PCL-R Facet 2</td>
<td>.65***</td>
<td>.79***</td>
</tr>
<tr>
<td>PCL-R Facet 3</td>
<td>.46**</td>
<td>-.08</td>
</tr>
<tr>
<td>PCL-R Facet 4</td>
<td>.44**</td>
<td>-.18</td>
</tr>
<tr>
<td>Antisocial PD</td>
<td>.23</td>
<td>-.05</td>
</tr>
<tr>
<td>Borderline PD</td>
<td>.06</td>
<td>-.13</td>
</tr>
<tr>
<td>Histrionic PD</td>
<td>.14</td>
<td>.17</td>
</tr>
<tr>
<td>Narcissistic PD</td>
<td>-.08</td>
<td>.23</td>
</tr>
</tbody>
</table>

*Note:* Spearman rho correlations between scores on the measures. The IPDE measures are all dimension scores.  
\[ p < .05, \text{ 2-tailed.} \quad p < .01, \text{ 2-tailed.} \quad p < .001, \text{ 2-tailed.} \]
3.3.2 Predicting Aggressive Behaviour

Reported in the second and third columns of Tables 3.2 and 3.3 is the number within the sample for which PCL-R and IPDE assessments were completed and available at the time of the study as well as mean (and standard deviation) scores. Two follow-up periods are presented in Tables 3.2 and 3.3, which report AUC values as well as their standard errors (columns 4 and 9) and 95% confidence intervals (columns 5 and 10), correlations between scores and (i) a dichotomous outcome variable (columns 6 and 11) as well as (ii) a frequency count of aggressive incidents (columns 8 and 13). Cohen’s $d$ effect sizes calculated using correlations between scores and dichotomous outcomes are also reported (columns 7 and 12). The dependent variables were incidents involving physical aggression resulting in damage to property (Table 3.2) and interpersonal physical aggression (Table 3.3). The most important findings for the measures are summarized in the following text.
Table 3.2. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the PCL-R and IPDE Dimensions for Incidents Involving Damage to Property.

<table>
<thead>
<tr>
<th>Measure</th>
<th>$n$</th>
<th>$\bar{x}$ (SD)</th>
<th>AUC ($SE$)</th>
<th>95% CI</th>
<th>Event$^a$ $r$</th>
<th>$d^b$</th>
<th>Freq$^c$ $r$</th>
<th>AUC ($SE$)</th>
<th>95% CI</th>
<th>Event$^a$ $r$</th>
<th>$d^b$</th>
<th>Freq$^c$ $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCL-R Total</td>
<td>43</td>
<td>27.6 (4.2)</td>
<td>.48 (.089)</td>
<td>.30-.65</td>
<td>-.03</td>
<td>-.07</td>
<td>.01</td>
<td>.55 (.120)</td>
<td>.31-.78</td>
<td>.01</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>Factor 1</td>
<td>43</td>
<td>11.0 (2.7)</td>
<td>.42 (.093)</td>
<td>.24-.60</td>
<td>-.10</td>
<td>-.21</td>
<td>-.15</td>
<td>.45 (.125)</td>
<td>.20-.69</td>
<td>-.06</td>
<td>-.14</td>
<td>-.10</td>
</tr>
<tr>
<td>Factor 2</td>
<td>43</td>
<td>14.4 (3.0)</td>
<td>.53 (.089)</td>
<td>.36-.71</td>
<td>.07</td>
<td>.15</td>
<td>.11</td>
<td>.65 (.126)</td>
<td>.40-.89</td>
<td>.15</td>
<td>.36</td>
<td>.19</td>
</tr>
<tr>
<td>Facet 1</td>
<td>43</td>
<td>4.7 (1.8)</td>
<td>.43 (.091)</td>
<td>.26-.61</td>
<td>-.12</td>
<td>-.25</td>
<td>-.13</td>
<td>.48 (.117)</td>
<td>.25-.71</td>
<td>-.04</td>
<td>-.09</td>
<td>-.04</td>
</tr>
<tr>
<td>Facet 2</td>
<td>43</td>
<td>6.4 (1.3)</td>
<td>.45 (.090)</td>
<td>.27-.62</td>
<td>-.04</td>
<td>-.09</td>
<td>-.09</td>
<td>.44 (.116)</td>
<td>.21-.67</td>
<td>-.07</td>
<td>-.17</td>
<td>-.11</td>
</tr>
<tr>
<td>Facet 3</td>
<td>41</td>
<td>6.7 (1.7)</td>
<td>.43 (.092)</td>
<td>.25-.61</td>
<td>-.01</td>
<td>-.03</td>
<td>-.06</td>
<td>.66 (.122)</td>
<td>.42-.90</td>
<td>.26</td>
<td>.65</td>
<td>.23</td>
</tr>
<tr>
<td>Facet 4</td>
<td>43</td>
<td>7.7 (1.9)</td>
<td>.61 (.092)</td>
<td>.43-.79</td>
<td>.15</td>
<td>.30</td>
<td>.23</td>
<td>.55 (.120)</td>
<td>.32-.79</td>
<td>.05</td>
<td>.12</td>
<td>.06</td>
</tr>
<tr>
<td>Antisocial PD</td>
<td>31</td>
<td>21.5 (8.4)</td>
<td>.49 (.106)</td>
<td>.28-.70</td>
<td>-.01</td>
<td>-.01</td>
<td>-.06</td>
<td>.42 (.152)</td>
<td>.12-.72</td>
<td>-.11</td>
<td>-.27</td>
<td>-.11</td>
</tr>
<tr>
<td>Borderline PD</td>
<td>31</td>
<td>9.6 (4.9)</td>
<td>.78*** (.084)</td>
<td>.61-.94</td>
<td>.47**</td>
<td>1.08</td>
<td>.42**</td>
<td>.73** (.111)</td>
<td>.51-.94</td>
<td>.32*</td>
<td>.82</td>
<td>.33*</td>
</tr>
<tr>
<td>Histrionic PD</td>
<td>30</td>
<td>3.4 (2.3)</td>
<td>.51 (.113)</td>
<td>.29-.74</td>
<td>.03</td>
<td>.05</td>
<td>-.05</td>
<td>.43 (.156)</td>
<td>.13-.74</td>
<td>-.08</td>
<td>-.19</td>
<td>-.10</td>
</tr>
<tr>
<td>Narcissistic PD</td>
<td>31</td>
<td>5.1 (4.8)</td>
<td>.50 (.110)</td>
<td>.29-.72</td>
<td>.05</td>
<td>.10</td>
<td>-.01</td>
<td>.48 (.150)</td>
<td>.19-.78</td>
<td>.06</td>
<td>.14</td>
<td>-.02</td>
</tr>
</tbody>
</table>
Note: \(^a\)Point biserial correlation between scores and dichotomous event occurrence. \(^b\)Cohen’s \(d\), converted from the point biserial correlation between scores and dichotomous event occurrence. \(^c\)Spearman rho correlation between scores and frequency count of incidents.

\(*p < .05\), 1-tailed. \(**p < .01\), 1-tailed. \(***p < .001\), 1-tailed.
Table 3.3. Mean Scores, AUC Values, Correlations, and Cohen’s $d$ for the PCL-R and IPDE Dimensions for Incidents Involving Interpersonal Physical Aggression.

<table>
<thead>
<tr>
<th>Measure</th>
<th>$n$</th>
<th>$\bar{x}$ (SD)</th>
<th>AUC (SE)</th>
<th>95% CI</th>
<th>Event$^a$ r</th>
<th>$d^b$</th>
<th>Freq$^c$ r</th>
<th>AUC (SE)</th>
<th>95% CI</th>
<th>Event$^a$ r</th>
<th>$d^b$</th>
<th>Freq$^c$ r</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCL-R Total</td>
<td>43</td>
<td>27.6 (4.2)</td>
<td>.49 (.092)</td>
<td>.31-67</td>
<td>.02</td>
<td>.05</td>
<td>.00</td>
<td>.65 (.102)</td>
<td>.45-85</td>
<td>.27</td>
<td>.60</td>
<td>.21</td>
</tr>
<tr>
<td>Factor 1</td>
<td>43</td>
<td>11.0 (2.7)</td>
<td>.55 (.098)</td>
<td>.36-74</td>
<td>.11</td>
<td>.23</td>
<td>.10</td>
<td>.74** (.096)</td>
<td>.55-92</td>
<td>.38*</td>
<td>.88</td>
<td>.32*</td>
</tr>
<tr>
<td>Factor 2</td>
<td>43</td>
<td>14.4 (3.0)</td>
<td>.41 (.086)</td>
<td>.24-58</td>
<td>-.08</td>
<td>-.15</td>
<td>-.15</td>
<td>.47 (.099)</td>
<td>.27-66</td>
<td>-.01</td>
<td>-.02</td>
<td>-.06</td>
</tr>
<tr>
<td>Facet 1</td>
<td>43</td>
<td>4.7 (1.8)</td>
<td>.52 (.095)</td>
<td>.33-70</td>
<td>.03</td>
<td>.05</td>
<td>.08</td>
<td>.72** (.089)</td>
<td>.54-89</td>
<td>.35*</td>
<td>.80</td>
<td>.32*</td>
</tr>
<tr>
<td>Facet 2</td>
<td>43</td>
<td>6.4 (1.3)</td>
<td>.62 (.092)</td>
<td>.44-80</td>
<td>.20</td>
<td>.42</td>
<td>.17</td>
<td>.71* (.109)</td>
<td>.50-92</td>
<td>.30*</td>
<td>.67</td>
<td>.27</td>
</tr>
<tr>
<td>Facet 3</td>
<td>41</td>
<td>6.7 (1.7)</td>
<td>.35 (.092)</td>
<td>.17-53</td>
<td>-.12</td>
<td>-.26</td>
<td>-.27*</td>
<td>.45 (.115)</td>
<td>.23-68</td>
<td>.07</td>
<td>.15</td>
<td>-.12</td>
</tr>
<tr>
<td>Facet 4</td>
<td>43</td>
<td>7.7 (1.9)</td>
<td>.51 (.093)</td>
<td>.33-70</td>
<td>-.02</td>
<td>-.03</td>
<td>.03</td>
<td>.46 (.103)</td>
<td>.26-67</td>
<td>-.07</td>
<td>-.15</td>
<td>-.03</td>
</tr>
<tr>
<td>Antisocial PD</td>
<td>31</td>
<td>21.5 (8.4)</td>
<td>.61 (.103)</td>
<td>.41-81</td>
<td>.23</td>
<td>.47</td>
<td>.23</td>
<td>.63 (.109)</td>
<td>.42-85</td>
<td>.25</td>
<td>.56</td>
<td>.19</td>
</tr>
<tr>
<td>Borderline PD</td>
<td>31</td>
<td>9.6 (4.9)</td>
<td>.68† (.096)</td>
<td>.50-87</td>
<td>.30*</td>
<td>.65</td>
<td>.35*</td>
<td>.57 (.115)</td>
<td>.35-79</td>
<td>.11</td>
<td>.23</td>
<td>-.02</td>
</tr>
<tr>
<td>Histrionic PD</td>
<td>30</td>
<td>3.4 (2.3)</td>
<td>.75** (.092)</td>
<td>.57-93</td>
<td>.42**</td>
<td>.96</td>
<td>.46**</td>
<td>.83*** (.080)</td>
<td>.67-99</td>
<td>.54**</td>
<td>1.39</td>
<td>.41*</td>
</tr>
<tr>
<td>Narcissistic PD</td>
<td>31</td>
<td>5.1 (4.8)</td>
<td>.56 (.111)</td>
<td>.34-78</td>
<td>.16</td>
<td>.33</td>
<td>.16</td>
<td>.64 (.136)</td>
<td>.37-91</td>
<td>.33*</td>
<td>.77</td>
<td>.17</td>
</tr>
</tbody>
</table>
Note: aPoint biserial correlation between scores and dichotomous event occurrence. bCohen’s d, converted from the point biserial correlation between scores and dichotomous event occurrence. cSpearman rho correlation between scores and frequency count of incidents.

*p < .05, 1-tailed. **p < .01, 1-tailed. ***p < .001, 1-tailed. †p < .06, 1-tailed.
Inconsistent with hypotheses concerning prediction of damage to property, PCL-R Total score as well as both Factor scores and all Facet scores failed to predict this outcome using either follow-up period, with AUC values all not significantly higher than would be expected by chance. However, the effect sizes for Factor 2 and Facet 3 (lifestyle features) fell in the small to medium range. Total and Factor 2 scores also failed to predict interpersonal physical aggression. In contrast, Factor 1 and both its constituent Facets, 1 and 2 were significantly predictive using the initial 12-month follow-up, as hypothesized, with AUCs of .74, .72, and .71, respectively, correlations of .38, .35, and .30, and corresponding Cohen’s $d$ values representing medium to large effect sizes. Correlations with frequency count for this outcome were also significant for Factor 1 and Facet 1. It is worth noting also the change in pattern among effect sizes according to outcome. For damage to property, the Cohen’s $d$ effect sizes for Factor 1 were small and negative while for Factor 2 they were small to medium and positive, indicating an association (albeit weak) between the lifestyle and antisocial features of psychopathy and the expression of aggression involving damage to property. This was reversed, with medium to large effect sizes for Factor 1 (and Facets 1 and 2) generally, when the outcome was interpersonal physical aggression, indicating an association between aggression directed towards other individuals and elevated scores in the interpersonal and affective domains of the PCL-R.

Turning to the IPDE dimension scores, the hypotheses for the outcome damage to property were only partially supported, with the score for Borderline PD shown to be statistically significant using either follow-up period (AUCs ≥ .73); correlations with the
dichotomous outcome variable and with frequency count were also significant ($r_s \geq .32$), representing large effect sizes.

For the outcome interpersonal physical aggression, there was, again, partial support for the hypotheses, with the dimension score for Histrionic PD shown to be a significant predictor using either follow-up period ($\text{AUCs} \geq .75$). The correlations with the dichotomous outcome variable and frequency count were also significant using either follow-up period ($r_s \geq .41$), representing large effect sizes. The dimension score for Borderline PD correlated significantly with the dichotomous variable and frequency count using the full follow-up period ($r_s \geq .30$), with a medium Cohen’s $d$ effect size. The dimension score for Narcissistic PD was correlated significantly with the dichotomous variable using the initial 12-month follow-up period ($r = .33$), again representing a medium effect size. As well, although neither the AUC values nor correlations for the Antisocial PD dimension score were significant, Cohen’s $d$ effect sizes were nevertheless roughly medium in size.

Two sets of partial correlation analyses were carried out to examine associations between outcomes and (i) the PCL-R, and (ii) IPDE dimension scores while controlling for the effect of other predictors. To limit the number of analyses, these analyses were restricted to the outcome interpersonal physical aggression and to associations with this outcome for PCL-R and IPDE variables identified as significant predictors on the basis of their point biserial correlations. Focusing first on PCL-R scores, partial correlations with interpersonal physical aggression occurring in the initial 12-month period controlling for Histrionic PD score ranged from .23 (for Total, Facet 1 and Facet 2) to .28 (for Factor 1), all statistically nonsignificant, representing effect sizes in the small to medium range.
The partial correlation between Facet 1 and interpersonal physical aggression controlling for Facet 2 was .25 (ns) and the partial correlation between Facet 2 and interpersonal physical aggression controlling for Facet 1 was .16 (ns), effect sizes again falling in the small to medium range.

For the IPDE dimension scores, analyses were limited to those examining Histrionic PD. Partial correlations with interpersonal physical aggression occurring in the initial 12-month period controlling for PCL-R Total, Factor 1, and Facets 1 and 2 scores ranged from .47 (\(p < .05\), 2-tailed, controlling for Factor 1) to .52 (\(p < .05\), 2-tailed, controlling for Facet 2). To determine the degree of overlap between PD dimension scores in the same cluster as Histrionic PD that might partially account for its strong association with interpersonal physical aggression, partial correlations controlling for dimension scores for Borderline PD, Antisocial PD, and Narcissistic PD were carried out also. These ranged from .35 (ns, controlling for Borderline and Antisocial PD separately) to .41 (\(p < .05\), 2-tailed, controlling for Narcissistic PD).

3.3.3 Predicting Repetitive Aggressive Behaviour

An additional series of ROC analyses were carried out to evaluate the accuracy of the measures in predicting two or more focal incidents over the full follow-up period. To limit the number of analyses, this series was also restricted to the outcome interpersonal physical aggression. Consistent with hypotheses, among the IPDE dimension scores, statistically significant values were found for Histrionic PD (AUC = .80, SE = .082, 95% CI = .64–.97), Antisocial PD (AUC = .72, SE = .102, 95% CI = .52–.92), and Borderline PD (AUC = .75, SE = .123, 95% CI = .51–.99).
3.3.4 Imminence of Incidents and Failure Rates

A series of Kaplan-Meier survival analyses were carried out to investigate whether higher scores on the measures were associated with faster time to an incident involving aggression. Again, to limit the number of analyses, this series were restricted to the outcome interpersonal physical aggression. The ROC analyses reported above indicated that scores on the PCL-R Factor 1 and Histrionic PD predicted interpersonal physical aggression so failure rates for these variables were investigated.

Scores on each of the measures were divided into four groups based on quartiles and the survival functions for groups plotted. For the PCL-R Factor 1 score, the difference between failure rates for the first (scores ≤ 9) and fourth quartile (scores > 13) was statistically significant (Log Rank test: $\chi^2 = 4.607, p = .032$), with the fourth quartile group showing the sharpest decline compared with the other groups (see Figure 3.1). No other comparisons revealed significant differences. At the 12-month mark, the cumulative proportion surviving in the first, second, third, and fourth quartile groups were 92%, 70%, 64%, and 37%, respectively.

For the Histrionic PD dimension score, the difference between failure rates for the first (scores ≤ 1.75) and fourth quartile (scores > 5) was statistically significant (Log Rank test: $\chi^2 = 4.108, p = .043$) as was the difference between failure rates for the second and third quartiles (Log Rank test: $\chi^2 = 4.094, p = .043$) (see Figure 3.2). At the 12-month mark, the cumulative proportion surviving in the first, second, third, and fourth quartile groups were 100%, 100%, 44%, and 42%, respectively. The pattern suggests that a division at the median value (3.50) provides the clearest distinction between groups’ failure rates.
Figure 3.1. Interpersonal physical aggression survival curves for PCL-R Factor 1 quartile groups.
Figure 3.2. Interpersonal physical aggression survival curves for Histrionic PD quartile groups.
3.4 Discussion

Consistent with the intake criteria for the DSPD service, one of the most immediately obvious findings of the study was that patients’ PCL-R scores were notably higher than those reported elsewhere in the literature. For example, the mean PCL-R Total score in the present sample was 62% higher than that reported by Gray et al. (2003) for a sample of 34 male medium security psychiatric inpatients in England, 30% higher than that reported by Reiss, Grubin, and Meux (1999) for a sample of 89 high security hospital inmates in England, 30% higher than that reported by de Vogel and de Ruiter (2005) for a sample of 42 male forensic psychiatric patients in The Netherlands, and 22% higher than that reported by Hare (2003) for a pooled sample of 1,246 male forensic psychiatric patients from a variety of jurisdictions. However, mean PCL-R score and range for the present sample was comparable with that reported for a sample of 48 offenders within a separate UK prison-based DSPD pilot program (Taylor, 2003: mean PCL-R Total = 26.3, range = 13–36). The most frequently diagnosed PDs in the sample, APD, BPD, and NPD, have been found by others to be among the most prevalent PD diagnoses in forensic psychiatric settings (Coid, Kahtan, Gault, & Jarman, 1999; Hildebrand & de Ruiter, 2004).

Despite the elevations in scores on the PCL-R in the present sample, it is interesting to note that the mean number of incidents of physical aggression involving damage to property and interpersonal physical aggression over the full follow-up period were actually lower than the respective mean number of incidents reported by Gray et al. (2003) over a 3-month period (the mean number of incidents involving verbal aggression was higher in the present sample than in that study but, of course, the follow-up period
was longer). One limitation of the present study is the reliance on the standard hospital incident forms for outcome data. Without additional sources to confirm the comprehensiveness of this single source it is possible that it reflects an underestimate of actual incidents of aggression occurring during the period. Of interest, however, Taylor (2003) reported a significantly lower number of violent incidents among offenders in a pilot DSPD program than expected based on pre-entry levels of violence. It is possible that entry into a regime such as characterizes the Peaks Unit, which includes a high level of security, high staff-to-patient ratios, clear and consistent structure, and group meetings to facilitate communication, might account for such discrepancies but this remains an empirical question. As others have observed, staff attention to early indicators of individual risk factors for physical aggression would be intended to reduce the number of incidents, which would, in turn, be expected to impact on the predictive validity of measures such as those examined in the present study (Hart, 1998; Taylor, 2003).

The objective for the present study was to determine if the assessment measures completed in the course of normal clinical practice would have a reasonable level of predictive validity in a forensic psychiatric inpatient sample for which scores were expected to have a restricted range that was skewed high. Overall, findings using the full follow-up period (which maximized the number of patients included but did not control for differing follow-up times among the patients) were not always matched by findings using the initial 12-month fixed follow-up period. It was for analyses using the latter that more findings were statistically significant despite the lower power associated by reductions in $n$. Nevertheless, the small sample size represents another notable limitation of the study. The results should be viewed with these limitations in mind.
3.4.1 PCL-R Performance

Only partial support for the study hypotheses were found with mixed results for the PCL-R, although the low statistical power of the study was clearly an issue. No component significantly predicted damage to property although Factor 2 and Facet 3 (the lifestyle/impulsive, irresponsible behavioural style facet) had effect sizes in the small to medium range (Cohen’s $d = .36$ and $d = .65$, respectively) using the initial 12-month follow-up, consistent with findings reported in meta-analyses for similar categories of outcome. In another small sample study ($N = 34$), Gray et al. (2003) reported that Total and Factor 2 scores significantly predicted this outcome (AUC = .76 and .87, respectively).

In terms of interpersonal physical aggression, in the present study PCL-R Factor 1 (and both its interpersonal and affective facets, 1 and 2) significantly predicted this outcome in the first 12 months (AUCs ≥ .71; $r ≥ .30$; Cohen’s $d ≥ .67$) and a survival analysis revealed that patients in the present study with Factor 1 scores in the highest quartile showed the fastest failure rate among the sample. However, PCL-R Total and Factor 2 scores were not significantly predictive of this outcome (although the correlation for the PCL-R Total score was $.27$ [$p = .061$] with a Cohen’s $d$ of $.60$). These findings contrast somewhat with those reported by Gray et al. (2003), but recall that in the recent meta-analysis by Walters et al. (2008) significant mean effect sizes were found for all four facets in predicting post-release violence. Possible reasons for the differences in magnitude of effect sizes between the results reported here and those of the Gray et al. study include the risk levels of the samples, the types of institution involved, the lengths of follow-up (3 months in the Gray et al. study, fixed 1 year and average 1. 5 year follow-ups in the present study), and differences in the way aggression was operationalized. In
regard to the latter, distinctions drawn in recent empirical work between aggression directed against staff versus other inpatients (McDermott et al., 2008) and between reactive versus instrumental aggression (Vitacco et al., 2009) will also be important in sorting out discrepancies in future studies. Such findings suggest that researchers attempt to identify optimal risk factors for distinct outcomes and revisit theoretical conceptualizations of how certain clinical features and personality traits might be more strongly associated with particular types and targets of inpatient aggression. Additional investigations that directly examine these and other possible moderators are clearly required.

Regardless, it is of considerable interest that the present study’s findings for Factor 1 are contrary to expectations that greater specification of psychopathy as a personality construct (for example, removing of items tapping antisociality when using the PCL-R) would lower its predictive validity (Skeem, Mulvey, & Grisso, 2003). Walters et al.’s (2008) meta-analysis found Facet 4 (antisocial) scores had incremental validity when combined with Facets 1 to 3 in predicting post-release recidivism in 11 of 11 analyses, but Facets 1, 2, and 3 added unique variance to recidivism predictions above that provided by Facet 4 in only 2 of 11 analyses. This pattern of results suggests that it is Facet 4 that underlies the superiority of Factor 2 and the PCL-R’s predictive validity more generally. Yet, others have reported data suggesting that Factor 1 may contribute unique explanatory power in prediction models. Heilbrun et al. (1998) found that of the two Factor scores only Factor 1 had incremental validity when added to age, race, and violence history in predicting total aggression during the initial two months of hospitalization in a US sample of 218 male forensic psychiatric patients.
It is possible that in psychiatric samples with a high mean PCL-R scores, such as found in the present sample, it is the predictive validity of the interpersonal and affective traits that are improved for interpersonal physical aggression. In samples with a range of scores with means closer to the published norms for the measure it may be the lifestyle and antisocial traits that are stronger predictors, particularly for more general, inclusive categories of aggression. Although the mean scores were high in the present sample, the distribution for Factor 1 approximated a normal curve (skewness value of -.283, standard error of skewness of .361) while the distribution for Factor 2 showed some departure from symmetry (skewness value of -.573, standard error of skewness of .361) with a longer left tail indicating more scores at the high end of the range. From a statistical point of view this could have been a factor in the differential performance of the two Factors.

Another possibility is that situational variables not considered in this study or any of the others discussed above served to moderate the performance of the measures (Gadon, Johnstone, & Cooke, 2006). A highly structured and continuously monitored environment such as is found on the Peaks Unit may have a differential impact on the rate of incidents of aggression that stem from features associated with the two PCL Factors. For example, perhaps antisocial lifestyle features of the individual enjoy more florid expression in less restrictive environments where a broader range of Factor 2 scores may then have greater predictive validity. In contrast, the expression of interpersonal and affective traits among those with highly elevated Factor 1 scores may be less susceptible to the influence of a structured and closely monitored environment (indeed, they might be more likely to be expressed among these individuals because of the social dynamics
between patients and with staff in such an environment) with the result that their predictive validity is improved.

Functional assessment of violent and aggressive behavior among samples such as the current one would help clarify this issue (Fox & Gable, 2004). Such a step would, in any case, be required in order to establish whether there is a functional link between symptoms of an individual’s personality disorder and acts of violence s/he perpetrates and so directly inform clinical efforts (Daffern et al., 2006; McGuire, 2008; Nolan et al., 2003). Regardless, it should be borne in mind that a change in environmental contingencies (e.g., a lower security institution or community-based setting) or a broadening of the sample studied so that individuals meeting DSPD criteria represent only a subset within a sample (e.g., Coid et al., 2007) would likely have a large impact on the predictive validity of the measure in group analyses.

### 3.4.2 IPDE Performance

Among the IPDE dimension scores, only Borderline PD significantly predicted damage to property (AUCs ≥ .73 and rs ≥ .32 for both full and 12-month follow-ups) with large effect sizes (Cohen’s $d \geq .82$). Taken together with the results for interpersonal physical aggression discussed below, these findings are consistent with what would be expected on the basis of diagnostic criteria for the disorder, notably emotional dysregulation, affective instability and difficulties controlling anger (Bradley, Conklin, & Westen, 2007; Howells, 2009).

For the outcome interpersonal physical aggression, AUC values for the Histrionic PD score were statistically significant for both follow-up periods (AUCs ≥ .75). The correlations with this outcome and frequency of such incidents were also significant ($rs$
and the Cohen’s $d$ effect sizes were large ($d_s \geq .96$). The only other dimension score that approached statistical significance in predicting interpersonal physical aggression was Borderline PD (AUC = .68, $p < .06$, $r_s \geq .30$ for the full follow-up period) with a medium effect size ($d = .65$). Consideration of these findings for Histrionic PD in terms of the broader literature is limited by the fact that there is a notable lack of research on its impact on any sort of significant outcome, including life satisfaction, relationship stability, health, and employment (Blagov, Fowler, & Lilienfeld, 2007). Nevertheless, the strong association between Histrionic PD and interpersonal physical aggression demonstrated here is consistent with the high rate of antisocial acts self-reported by histrionic men (Luisada, Peele, & Pittard, 1974) and its co-occurrence with Antisocial PD (Lilienfeld, VanValkenburg, Larntz, & Akiskal, 1986), the latter being firmly associated with violence (de Brito & Hodgins, 2009; Wormith, Olver et al., 2007).

Contrary to study hypotheses for Antisocial and Narcissistic PD dimension scores, neither significantly predict interpersonal physical aggression according to the AUC values (when the outcome was a single incident), although the fact that 73% of the sample was diagnosed with Antisocial PD likely impacted on the predictive accuracy of its dimension score. In any case, the effect size was medium as was the effect size for the Narcissistic PD dimension score. As with Histrionic PD, there is a profound empirical neglect of Narcissistic PD in the broader literature (Blashfield & Intoccia, 2000) with the vast majority of the published work being clinical in nature (Levy, Reynoso, Wasserman, & Clarkin, 2007). Yet such work does provide grounds to have expected a stronger association with aggression in the present study (Logan, 2009), and the possibility that
subtypes of Narcissistic PD (Ronningstam, 2005) might be differentially related to outcomes including types of aggression merits attention in future work.

For the outcome of repetitive interpersonal physical aggression (i.e., 2 or more incidents), the Antisocial PD dimension score was significantly predictive as were the Histrionic and Borderline PD dimension scores, with large effect sizes, suggesting that elevations on these PD dimension scores have clinical utility in identifying individuals at risk of repeated acts of interpersonal violence.

These findings are generally consistent with the factor-based associations with violent criminal history and self-reported anger/hostility discussed earlier (Howard et al., 2008; Ullrich & Marneros, 2004, 2007). Indeed, Howard et al.’s first order factors, Antisocial, Histrionic-Narcissistic, and Borderline, had high loadings (.84, .52, and .40, respectively) on their second-order Antisocial factor, which itself had a range of significant associations with indices of anger, impulsivity, and criminal history. Not only was partial support found for the hypotheses concerning the predictive validity of PD dimensions scores in the current study, as was found for the PCL-R Factor 1 scores, survival analysis demonstrated that higher Histrionic PD dimension scores were associated with imminence of interpersonal physical aggression.

Although the direct treatment implications of these personality trait data for high risk, personality disordered patients are less clear than are findings concerning structured clinical risk assessment instruments comprised of dynamic factors (see chapter 2), such traits do provide a starting point from which to theorize and then investigate levels of explanation (Blackburn, 2007a; Duggan & Howard, 2009). Further, at least tentative speculations can be offered in regard to the potential utility of these traits as
intraindividual risk indicators (Douglas & Skeem, 2005) over the medium to long term. Turner and Dudek (1997) have suggested that repeated PD ratings over time that index variation in patients’ functioning afford a potential focus for assessment of treatment-related change, although triangulating evidence of such change will undoubtedly present difficulties (Duggan, 2004). Certainly dimensional representations of PDs have empirical support as well as theoretical and clinical advantages over categorical approaches (Livesley, 2001; Widiger & Frances, 2002), and the evidence indicates that personality features can and do change over time (Blackburn, 2000; Sanislow & McGlashan, 1998; Tyrer et al., 2007; Warner et al., 2004), with dimensional approaches potentially offering a more sensitive means by which to assess that change than do categorical approaches. Of course, treatment readiness and motivation versus resistance among personality disordered forensic patients (Howells & Day, 2007; Tyrer, Mitchard, Methuen, & Ranger 2003; Wormith & Olver, 2002) provide considerable challenges for clinicians, and require that researchers examine these variables as potential mediators and moderators in their efforts to demonstrate that change can be attributed to treatment (Langton, 2007).

3.4.3 Functional Link Between Personality Traits and Risk of Future Aggression

These findings are also of relevance in considering the explicit DSPD emphasis on a ‘functional link’ between personality disorder and risk of future harm to others (criterion three for entry into the service). Establishing an empirical basis for such a link would mean it need not be presumed only on the basis of individual case formulation using clinical judgment. But it is not clear from the existing literature on the DSPD initiative how this functional link is intended to be conceptualised or operationalized. Duggan and Howard (2009) suggest that this criterion be interpreted as “implying a
‘causal connection’ between the severity of the PD [personality disorder] and the resultant increased risk of violent behaviour. Put simply, it is the ‘severity of the PD’ that ‘causes’ the individual to ‘behave violently’.” (p. 20). They discuss four general conditions necessary to determine whether a causal functional relation exists between variables (Haynes, 1992): (i) covariation between the variables, (ii) temporal precedence of the causal variable, (iii) exclusion of an alternative explanation for the relationship, and (iv) establishment of a logical connection between the variables.

Importantly, the link is specified as a criterion for entry into the service with no comment about whether the link (its maintenance or severance) be considered for intervention planning or security changes. It is not clear that the link would remain given that concerted efforts to break it and prevent further violence through management and treatment would be expected. This makes the question of whether the link remains intact for some disorders/traits (assuming the link was established pre-admission) of considerable interest with regard to ongoing planning for patients and the service more broadly. It is well-established that prior criminal and violent behaviour predicts future criminal and violent behaviour (Bonta, Law, & Hanson, 1998). As such, inferring a functional link with risk of re-offending for personality disorders at the group level of analysis could be based on the prospective prediction of inpatient aggression using personality traits, but only if the conceptualization and assessment of these traits is not confounded by violence/offence-related symptoms. Such an investigation could serve to address the first two of Haynes’ four criteria discussed by Duggan and Howard (2009). Use of IPDE dimension scores (excepting those for Antisocial PD) in a prospective study design are arguably appropriate for this purpose. Although scores on Factor 1 on the
PCL-R (the interpersonal and affective aspects of psychopathy) might afford a less compelling test of the functional link given that these items are scored, in part, on the basis of offence-related behaviour (Widiger, 2006) as well as patients’ thoughts and affective responses to their antisocial behaviours, arguably they might also be considered relevant to this issue (Cooke et al., 2007).

To examine such a purported ‘link’ from an empirical basis would require operationalizing the necessary and sufficient conditions to test Haynes’ (1992) four criteria. The present study was not undertaken for this purpose (and not all interpersonal physical aggression incidents necessarily resulted in serious physical or psychological harm) but the findings allow some comments in this regard. In the sample, Cluster B personality traits/dimensions that include no history of conduct disorder or criminality among their scoring criteria were prospectively predictive of interpersonal physical aggression, providing preliminary support for Haynes’ ‘covariation’ and ‘temporal precedence’ criteria. Further, the partial correlation data indicated that Histrionic PD dimension scores remained significantly correlated (and, perhaps more importantly, still had medium effect sizes) with interpersonal physical aggression after controlling for PCL-R Total, Factor and Facet scores. The latter represent important constructs (albeit a few among many) that would otherwise represent an alternative explanation for the association (Haynes’ third criteria). As an avenue for further research, it would be important to examine other personality disorders and determine if there are core traits for each of the clusters for which the link can be demonstrated (among both pre-admission candidates for the DSPD service as well as admitted inpatients). Of course, replication of
these findings will be required with designs utilizing multivariate causal modeling if
more is to be made of the concept of a functional link within this population.

### 3.4.4 Overlap Between Constructs and Implications for Prediction Models

A few last comments are warranted concerning associations between measures
and implications in terms of the incremental predictive validity that PD dimension scores
might have when added to the PCL-R in prediction models. There was generally a
reasonably high level of concordance between discrete PD dimension scores within
Cluster B but there was a notable lack of association between Factor 1 and 2 of the PCL-
R, contrary to findings reported elsewhere that indicate a moderate positive association
(e.g., Hare, 2003). This finding for the sample clearly requires further study as part of
efforts to evaluate uses of the PCL-R in clinical settings (as opposed to research
initiatives which typically have greater resources and quality controls in place) with
larger samples composed entirely of high risk personality disordered forensic psychiatric
patients.

In terms of inter-measure association, there is currently a dearth of studies
examining IPDE dimension scores and the PCL-R. In one important exception,
Blackburn (2007b) reported medium to large correlations between both PCL-R Factors
and Cluster B IPDE dimension scores. Others have reported few Factor 1 correlates with
self-report measures of PDs (Hart, Forth, & Hare, 1991; Shine & Hobson, 1997). In the
present sample there was limited consistency with what might have been expected based
on conceptual overlaps. The correlation between scores on Facet 1 and Narcissistic PD
dimension scores was .25, reflecting the interpersonally exploitative features and
grandiose sense of self worth shared in common. The Antisocial PD dimension score
was not correlated with PCL-R Facet 2, despite common features between the two such as lack of empathy or sense of responsibility. However, expected associations with Factor 2 and Facet 4 (antisociality) were seen (significant correlations of .46 and .49, respectively), reflecting the insufficient planning/impulsivity and difficulty controlling temper/poor behavioral controls in common.

The distribution of PCL-R Factor scores in this specially selected sample and the low statistical power associated with its size (a limitation that is certainly acknowledged) might go some way towards explaining the inconsistencies with Blackburn’s (2007b) correlation patterns. An additional limitation of the study, the absence of inter-rater reliability data, could also be relevant here. Within a separate DSPD unit, the IMPALOX group (2007) reported inter-rater reliability indices for the PCL-R and IPDE that were only fair. However, the indices of predictive and convergent validity reported here for the measures, used in a clinical rather than research context, imply satisfactory reliability and provides some indication of how they perform in a real-world application.

Based on his review of the pertinent literature, Blackburn (2007b) views the array of findings as “more consistent with the proposal that psychopathy be viewed as a superordinate dimension of personality deviation pervading several PDs” (p. 12, italics in original). His strong arguments notwithstanding, the present study’s findings concerning prediction accuracy taken together with the partial correlation data are of potential practical significance because they suggest that scores on certain IPDE dimensions and Factor 1 of the PCL-R might have incremental predictive validity within a comprehensive risk assessment/management approach (i.e., a patient’s PCL-R and IPDE Histrionic dimension score might both provide unique information pertinent to risk assessment and
management), at least for high risk, personality disordered forensic psychiatric inpatients.

The complex question that follows is whether or not:

(i) positive effects can be accrued by targeting in treatment for these patients their maladaptive styles of self-representation, (hostile) attribution biases, social skills deficits and other behavioural expressions of features characterizing Histrionic, Borderline and Antisocial personality features as criminogenic needs (Andrews, Dowden, & Rettinger, 2001; Duggan, 2008; Harkness & Lilienfeld, 1997; Livesley, 2007)

(ii) while tailoring the goals, format, and delivery of service components in light of the specific responsivity implications for treatment of elevated PCL-R Factor 1 scores (Andrews & Bonta, 2006; Wong et al., 2007; Wong & Hare, 2005; Wormith, 2007).

Addressing this question will require replication of the present findings with larger samples, more refined conceptual and theoretical modeling, and concerted clinical efforts. Certainly, the use of the IPDE and PCL-R within the DSPD program demands further investigation if either is to have clear implications for clinical practice after admission.
CHAPTER 4: General Discussion

In this brief, concluding chapter an overview of the main findings from the two empirical studies are first presented. Consideration is then given to a number of the limitations associated with the project and how these can inform future investigations of this kind. Finally some general implications for research and practice are explored.

4.1 Overview of Findings in Light of Research Objectives

The research project described in the previous chapters was undertaken as a component of the service evaluation project at one of the four national DSPD sites. The principle objective was to evaluate the criterion predictive validity of four risk assessment tools and two measures of personality traits with a sample representing 15% of the DSPD population. The research question was whether or not they would perform in the DSPD setting as they have been shown to do for other populations of offenders and psychiatric patients.

The risk assessment tools were designed to predict types of future violence (and criminal recidivism) and have been validated with a variety of populations. For some of them (the HCR-20), preliminary findings reported elsewhere suggested they would predict types of aggression within a high security forensic psychiatric setting although for others (the VRS) published studies have reported mixed results in secure psychiatric settings, and for others (the RM 2000 and the Static-99) there are no existing data concerning the prediction of institutional aggression. Despite this, evaluation research for all of them strongly indicates that these tools include empirically established risk factors for future violence permitting assessors to meaningfully rank individuals in terms of their likelihood of engaging in future aggressive behaviour. On this basis, all four
tools were hypothesized to predict two types of institutional aggression: damage to property and interpersonal physical aggression.

Turning to the two personality measures, there are considerable differences in the scope and depth of the published research on the predictive validity between them. The sizable literature on the association between scores on the PCL-R (and its Factors and Facets) and future violence, including types of institutional aggression provided a solid foundation on which to base hypotheses in the current research concerning the accurate prediction of the two types of institutional aggression. Although IPDE dimension scores had not been evaluated in terms of association with future violence prior to the research reported here, factor analytic studies involving psychiatric patients and offender samples strongly suggests that Cluster B traits are associated with a history of antisocial and aggressive acts. Despite the relative dearth of research on long-term outcomes (both in terms of crime/violence and other important domains of life) for individuals with personality disorders, there is sufficient empirical evidence and clinical conceptualizations to support the hypotheses in the current research that IPDE Cluster B dimension scores would predict the two types of institutional aggression.

Results provided partial support for a number of the hypotheses. Among the risk assessment tools, the HCR-20 Total, Clinical, and Risk Management scales as well as the SFRJ showed at least moderate levels of predictive accuracy for both outcomes for at least one of the two follow-up periods. The SFRJ also predicted repetitive interpersonal physical aggression and scores on the Risk Management scale were associated with imminence of interpersonal physical aggression. However, the Historical scale (comprised of static risk factors based on historical information) failed to predict either
outcome. These findings are generally consistent with results reported elsewhere. Of note, however, the HCR-20 SFRJ has not been examined in published studies to date with forensic psychiatric inpatients; the positive findings are thus promising but must be considered preliminary only. In contrast, performances for the VRS (Total score and Static and Dynamic Factor scales), the RM 2000 and the Static-99 were all less consistent and generally weaker. Total scores for these tools did predict damage to property in the first year following admission but no total or scale scores predicted interpersonal physical aggression. Effect sizes varied and those few in the medium range do suggest that additional research with larger samples is required before conclusions are drawn about the predictive utility of these tools within the DSPD service. This is particularly the case for the VRS Dynamic Factor scale, which has considerable potential as means to assess ranges in risk over time and so inform treatment planning.

Partial support was also obtained for the hypotheses made regarding the personality measures. No component of the PCL-R predicted damage to property but Factor 1 and its constituent Facets predicted interpersonal physical aggression within the first year after admission. Factor 1 scores were also associated with imminence of interpersonal physical aggression. PCL-R Total score had a medium effect size but fell short of statistical significance in predicting interpersonal physical aggression. In contrast to Factor 1, Factor 2 showed no association with either outcome. Various explanations were considered for this in the previous chapter, all of which represent important avenues for further empirical enquiry.

Mixed results were also found for the IPDE Cluster B dimension scores. Borderline Personality Disorder predicted damage to property but fell short of statistical
significance for the interpersonal physical aggression outcome. In contrast, Histrionic Personality Disorder predicted this outcome and, along with Antisocial and Borderline Personality Disorders, predicted repetitive interpersonal physical aggression also. The dimension score for Histrionic Personality Disorder was also associated with imminence of interpersonal physical aggression. A preliminary attempt to examine associations with this outcome while controlling for the effects of other predictors revealed a statistically significant partial correlation for Histrionic Personality Disorder that remained whether the controlled predictor was PCL-R Factor 1 or other Cluster B dimension scores.

4.2 Revisiting Limitations of the Project

As has been acknowledged in the preceding chapters, the sample size available for the research reported was very small. Although it represents what is arguably a reasonable proportion of the total DSPD population, such a small sample permits only tentative conclusions be drawn from the results and limits the overall generalizability of the two studies. As an exercise in hypothesis generation, there is much to consider from these data, as reviewed in the preceding chapter discussions. In clinical practice, the lack of empirical data to inform the provision of services alongside efforts to manage (and reduce) risk of violence in a population explicitly selected for its high needs and high risk means preliminary work of this kind can be useful. But replication with a larger sample from the population is imperative and, on the basis of these findings, it would seem that this set of risk assessment tools and personality measures should not be relied upon as the only means by which to assess patients’ risk for institutional violence. Thus, expanding the range of instruments examined for this purpose with a larger sample from the DSPD population is clearly indicated also.
Also associated with the small sample size problem is the low statistical power available to test hypotheses. A larger sample would have also meant that a variety of additional analyses could have been supported. A high priority in terms of generating and evaluating theoretical models of violence prediction as well as informing clinical practice would be analyses of incremental validity (Garb, 2003; Hunsley & Meyer, 2003). Certainly dynamic risk factors can be assessed reliably and effectively addressed by interventions (Douglas & Skeem, 2005; Douglas, Nicholls, & Brink, 2009; Douglas, Webster, Hart, Eaves, & Ogloff, 2001). Within the DSPD setting, it is important to determine whether dynamic variables, with their considerable potential to inform interventions targeting the clinical needs and situational factors associated with higher risk of aggression, account for unique variance in statistical models in which static factors (or at least variables less amenable to change over weeks, months, or even years such as personality traits) are entered first. Underlying investigations of this kind are questions concerning the elements of a comprehensive risk assessment, how to recognize information that adds something to the formulation and resulting intervention plan, how to avoid giving weight to redundant or misleading information, how to gauge change in risk, and how to balance the tension between optimizing prediction accuracy on the basis of certain risk factors (some of which may be quite distal from the adverse event and not amenable to change) and preventing the event that one is attempting to predict by addressing other risk factors (that are more proximal but perhaps lack incremental validity).

Another limitation of the two studies reported is the use of aggression category outcomes (damage to property and interpersonal physical aggression) that did not make
further distinctions between types of aggression. A growing body of work with forensic patients supports the parsing of an inclusive category into instrumental aggression (that is, planned, goal directed aggression, which might be best predicted by a particular constellation of personality traits) and aggression that is reactive in nature (aggression resulting from provocation and arousal to hostility, which might be optimally predicted by proxies for poor behavioural control) (Cornell et al., 1996; Fontaine, 2007; Kockler, Stanford, Nelson, Meloy, & Stanford, 2006). In a sample of 152 male forensic inpatients drawn from minimum, medium, and maximum security levels within a state mental health facility, Vitacco et al. (2009) found that anger scales (assessed using the staff-completed Ward Anger Rating Scale: Novaco, 1994a) and symptom scales for mental illness (assessed using the staff-completed BPRS: Overall & Gorham, 1962) predicted reactive aggression. In contrast, and of direct relevance to the studies reported in the preceding chapters, instrumental aggression was predicted by the Interpersonal, Affective, and Antisocial Facets of the PCL:SV (Hart et al., 1995) while none of these Facets predicted reactive aggression and only the Antisocial Facet predicted a category combining both types of aggression. This differential pattern of results suggests that various clinical features and personality traits predict different kinds of aggression, patterns that cannot be discerned when broader, general categories of aggression are employed.

At present, a notable feature of the literature is the significant number of variations in the way aggression has been operationalized, which likely goes someway towards explaining the variations in results (Guy et al., 2005). Greater standardization as well as the drawing of finer distinctions among the types of aggression represents an
avenue for further research that has both theoretical and applied significance. From a research perspective, however, such effort does bring with it the statistical problem posed by the generally low base rates for inpatient aggression (Rogers & Shuman, 2005). Although a base rate of 39% for interpersonal physical aggression available for the analyses reported in the present studies is higher than has been reported a number of studies of this kind (for example, Vitacco et al., 2009), additional divisions of aggression outcomes would necessary lower the rate for each. This is a widely recognized problem when attempting to predict rare events (Martin & Terris, 1991, Quinsey, Harris, Rice, & Cormier, 2006). Nevertheless, the potential value of identifying predictors of different types of aggression for both theoretical and clinical work is substantial and so future research should certainly incorporate such distinctions.

Another consideration in efforts to better operationalize types of aggression in studies of this kind concerns the target of the aggression. Using a sample comprised of both male and female forensic inpatients, McDermott et al. (2008) found distinct differences in prediction performance according to whether patients’ aggression was directed towards staff or other patients. Symptoms of mental illness (assessed by staff using the hostility, positive, negative, depressive symptoms scales of the BPRS [Overall & Gorham, 1962]) predicted aggression towards other patients but not staff. In contrast, anger (assessed using a self-report instrument, the Novaco Anger Scale and Provocation Inventory [Novaco, 1994b] predicted aggression towards staff but not other patients. Of direct relevance to the present studies, Total score on the HCR-20 as well as the Clinical and Risk Management scales in McDermott et al.’s sample predicted a category combining aggression towards both staff and patients, but indices were higher for patient-
directed aggression while only Total and Historical scale scores were significantly
predictive of staff-directed aggression. As well, only Factor 2 of the PCL-R predicted
both types of aggression while the single other score reaching statistical significance for
patient-directed aggression was that for Facet 1. Although caution is required in
extrapolating from McDermott et al.’s data because of the mixed sample and the
relatively low mean PCL-R score ($M = 16$), the results have implications for the present
studies. It is possible that using a combined staff-and-fellow-patients aggression outcome
may have obscured differential relationships between the predictors and outcomes in the
studies reported in the preceding chapters, representing another important limitation of
the work. Such distinctions certainly merit attention in future investigations.

4.3 Some General Implications for Research and Practice

Consideration of the above limitations generates additional general implications
for further research and practice. As with any assessment within a clinical context, the
referral question should be central to the selection of procedures and measures to be used.
As has already been mentioned, the assessment battery used to determine entry into the
DSPD service includes tools that provide a means to monitor fluctuation in dynamic risk
over time and thus a possible mechanism to assist in decisions regarding both the timing
of interventions as well as discharge from the service. But questions about construct and
predictive validity remain to be addressed. It would be important to determine whether
these tools can assist in identifying when a patient, perhaps newly admitted to a ward,
poses an acute risk of violence towards others over hours or days or whether the
timeframe in which such violence can be predicted to occur is in terms of months
Tools for assessing acute risk will be required if the latter is found to be the case.

Many of the variables comprising these personality measures and risk assessment tools overlap (in construct or content). Nevertheless, the timeframe over which they show predictive validity likely varies as a function of differences in the ways they are appraised (for example, as a symptom present in the moment, such as elevated anger), or a more general expression of an underlying proclivity (for example a hostile attribution bias or pattern of poorly controlled anger across situations). In addition to investigating this, future research will need to determine the timeframe over which administrations of dynamic risk assessments can optimally detect clinically meaningful change and whether the changes are associated with changes in (particular types of) incidents of aggression over specified timeframes.

Another question, concerning a different aspect of construct validity, would be whether these changes correspond to changes in theoretically driven, treatment-specific targets. For example, do changes in insight (as operationalized by items on the Clinical scale of the HCR-20 and the Dynamic Factors scale of the VRS) correlate with changes in objective ratings made independently pre- to post-treatment by treatment providers of, for example, quality of behavioural chain analyses completed by those patients participating in Dialectical Behavior Therapy (Linehan, 1993; McCann, Ivanoff, Schmidt, & Beach, 2007)? Do changes in negative attitudes or cognitive distortions (additional HCR-20 and VRS items) correlate with changes in, for example, scores on the Psychological Inventory of Criminal Thinking (Walters, 2002) self-reported by patients participating in psychosocial interventions (Beck, Freeman, Davis, & Associates, 2004;
McGuire, 2008)? Does improved management of symptoms of mental illness (aspects of which are tapped by HCR-20 and VRS items) correlate with changes in, for example, scores on the BPRS (Overall & Gorham, 1962) in patients receiving psychopharmacological treatment (Links, Boggild, & Sarin, 2001)?

Further, it seems important to determine whether the criterion-prediction validity of these item sets or scale scores compare favourably with that of theoretically driven, treatment-specific objective ratings or scores on self-report measures. Of course, tools with good criterion-prediction validity (such as the PCL-R or the HCR-20) need to offer advantages beyond other tests or measures routinely used in forensic inpatient contexts that might have comparable validity in predicting violence. Trade-offs against criterion-prediction validity that require consideration include the level of training and time required to administer such a tool and its clinical utility. When a clinician rating tool or self-report measure can achieve a comparable level of prediction accuracy and similarly identify intervention targets (for example, the BPRS; see Gray et al. [2003], McDermott et al. [2008], and Vitacco et al., [2009]), additional comparisons of the effectiveness of the tools in terms of cost and impact on treatment should be considered (Yates & Taub, 2003). Under such circumstances, it behooves mental health professionals and direct care providers to be explicit about why certain tools are employed as well as how they inform risk management plans and contribute to beneficial treatment outcome in a coordinated and complimentary fashion (Nelson-Gray, 2003). On the basis of empirical work addressing these issues, it should be possible to shape policy and practice in order to develop optimal strategies for assessing and managing ongoing risk of violent behaviour.
and inform service delivery within explicitly examined systems, work representing the ‘fourth generation’ in assessment research of this kind (Andrews et al., 2006).

Additional research questions that deserve attention in future research of this kind concerns the role of moderators (Holmbeck, 1997), akin to the ‘interactionist questions’ being asked in areas of correctional psychology (Clements & McLearen, 2003). The results of the studies reported in chapters 2 and 3 suggest that population and setting might have a significant impact on the predictive validity of personality measures and risk assessment tools (Eden, 2006; Guy et al., 2005). Of course, the data reported in the preceding chapters are only preliminary and in need of replication with other samples from the DSPD population as well as other high risk secure populations. For example, in terms of settings, for the DSPD service it will be necessary to determine whether there are differences in the tools’ performance between patient samples housed within facilities in the prison or mental health systems. Another consideration is the type of treatment provided to address criminogenic needs and the fidelity with which such programs are implemented in these settings (McGuire, 2008; Wormith, Althouse et al., 2007).

Consider also participation in interventions (Smith & Gendreau, 2007)\(^6\). In research with sexual offenders, predictive validity for validated risk assessment tools such as the Static-99 have been found to vary as a function of treatment status (i.e., offender refused to participate in a sex offender-specific cognitive behavioural treatment program, dropped out of the program, or completed the program; Haag, 2005; Langton, Barbaree, Hansen, Harkins, & Peacock, 2007). Of course, the design of such studies does not permit inferences about the effects of intervention efforts because unmeasured variables could

\(^6\) Incorporation of service level information represents a promising avenue for further work in evaluation research in this area although presently there are few tools that could be considered representative of this fourth generation approach to risk assessment (see Andrews et al., 2006).
account for the differences in recidivism rates among the groups. But such results do underscore the need to consider heterogeneity in the sample under investigation.

Along with setting and treatment involvement, psychopathy and Cluster B traits would be high on the list of candidate moderator variables to be looked at, another means by which populations and samples might be more meaningfully parsed in research efforts to better determine optimal applications for the tools in applied work. That said, it must be acknowledged that current diagnostic criterion sets for personality disorders are themselves far from adequate in ensuring homogeneity; indeed, the low levels of reliability and validity in so many respects seems to demand comprehensive overhaul in future work (Cloninger, 2007; Fowler, O’Donohue, & Lilienfeld, 2007; Tyrer et al., 2007; Widiger, 2007).

Nevertheless, the potential clinical utility of incorporating personality constructs into risk assessment and management models (beyond simple inclusion as one more variable in professional guidelines or as a weighted predictor in an actuarial scheme such as how the PCL-R has been used in the HCR-20 and Violence Risk Appraisal Guide [Harris et al., 1993]) is considerable. The PCL-R has been shown to function as both a risk factor (predicting violence) and a responsivity factor (to use Andrews and Bonta’s [2006] terminology), adversely affecting treatment outcomes (e.g., Hobson, Shine, & Roberts, 2000), and so should prompt modification of service content and delivery. However, because it was developed to assess psychopathic traits over the lifespan (and is therefore expected to remain stable over time), it does not afford a sensitive index of clinical constructs that might respond to intervention.
In contrast, IPDE dimension scores for Cluster B traits (or, indeed, alternative dimensional models; Widiger & Simonsen, 2005) might well function in all three capacities; predicting outcomes such as violence, moderating response to interventions, and providing targets for specifically designed treatment (intended to shape characteristic adaptations arising from the traits and not necessarily the traits themselves; Harkness & Lilienfeld, 1997). But confirmation of such functions for the IPDE or any personality disorder model or assessment approach remains to be empirically demonstrated at this point in time (Duggan et al., 2007). Further, detecting change with the IPDE would necessarily require extended follow-up periods and some variation in setting given the scoring instructions and the general diagnostic criteria for a personality disorder. These include “an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual’s culture… [that] is stable and of long duration… [and that] is inflexible and pervasive across a broad range of personal and social situations.” (APA, 2000, p. 689, italics added).

As a closing observation, it is still the case that despite the considerable progress in the field in the last two decades (Monahan, 1996; Hanson, 2009), an empirically supported, theoretically coherent approach to combining information from different personality measures, risk assessment tools, and treatment progress indices is still to be fully realized. Statistical models have not yet determined how to combine scores for optimizing prediction accuracy or to reconcile differences in risk level between different actuarial risk assessment instruments (Barbaree, Langton, & Peacock, 2006; Seto, 2005) but professional guidelines (Douglas & Kropp, 2002; Webster, Hucker, & Bloom, 2002) and the treatment literature from correctional psychology (Duggan, 2008; Wormith,
Althouse et al., 2007) provide direction both for service providers and researchers at this stage. In the interim, clinicians are well advised to attend to the distinctions between prediction and prevention (Heilbrun, 1997) and remain abreast of advances in both the empirical literature and recommended best practices (McNeil, 2009; Melton, Petrila, Poythress, & Slobogin, 2007; Monahan, 2003).
REFERENCES


disordered. In J. Monahan & H. J. Steadman (Eds.), *Violence and mental
disorder: Developments in risk assessment* (pp. 21-59). Chicago: University of
Chicago Press.

Angeles, CA: Western Psychological Services.

An instrument to assess risk for imminent aggression in psychiatric inpatients.

*Psychological Reports, 10*, 799-812.

coercion in treatment: A meta-analysis of effectiveness. *Criminal Justice and
Behavior, 35*, 1109-1135.

approach to modeling the structure of the structure of the Psychopathy Checklist-
Revised. *Journal of Personality Disorders, 21*, 118-141.

Peay, J. (2007). Mentally disordered offenders, mental health, and crime. In M. Maguire,
R. Morgan, & R. Reiner (Eds.), *The Oxford Handbook of Criminology* (4th


Monahan & H. J. Steadman (Eds.), *Violence and mental disorder: Developments

Wong, S., & Gordon, A. (2000). *Violence Risk Scale*. Saskatchewan, Canada:
Authors.

Risk Scale: A treatment friendly violence risk assessment tool. *Psychology,
Public Policy, and Law, 12*, 279-309.

prone forensic clients: an integrated approach. *British Journal of Psychiatry, 190
(suppl. 49)*, s66-s74.

Program*. Toronto: Multi-Health Systems.

health professionals* (pp. 83-120). New York: Springer.

H. L. (2005). Reintegrating seriously violent and personality disordered
offenders from a supermaximum security institution into the general offender
population. *International Journal of Offender Therapy and Comparative
Criminology, 49*, 362-375.


# APPENDIX 1

## Glossary of Terms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>First/Main mention</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>BPRS</td>
<td>Brief Psychiatric Rating Scale</td>
</tr>
<tr>
<td>DSM</td>
<td>Diagnostic and Statistical Manual</td>
</tr>
<tr>
<td>DSPD</td>
<td>Dangerous and Severe Personality Disorder</td>
</tr>
<tr>
<td>HCR-20</td>
<td>Historical Clinical Risk – 20</td>
</tr>
<tr>
<td>HCR-20 SFRJ</td>
<td>Structured Final Risk Judgment using the HCR-20</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Disorders</td>
</tr>
<tr>
<td>IPDE</td>
<td>International Personality Disorder Examination</td>
</tr>
<tr>
<td>PCL-R</td>
<td>Psychopathy Checklist-Revised</td>
</tr>
<tr>
<td>PCL:SV</td>
<td>Psychopathy Checklist: Screening Version</td>
</tr>
<tr>
<td>PD</td>
<td>Personality Disorder</td>
</tr>
<tr>
<td>RM-2000/S</td>
<td>Risk Matrix 2000/Sexual scale</td>
</tr>
<tr>
<td>RM-2000/V</td>
<td>Risk Matrix 2000/Violence scale</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>VRS</td>
<td>Violence Risk Scale</td>
</tr>
</tbody>
</table>
APPENDIX 2
Participant Information Sheet

**Study Title:** The retention of routine clinical information on a data base for research purposes from patients admitted to the Peaks unit

**Name of Principal Researcher:**
Professor Kevin Howells, Head of Peaks Academic and Research Unit

**Invitation to participate**
You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

*Thank you for reading this.*

**What this research is about?**
At the present time, there is a great deal of interest in the assessment and treatment of personality disorder. This is an area of mental health that has been neglected for a long time, so there is a great deal that we need to find out in order to use treatments most effectively.

In order to increase our knowledge, it is important to examine how individuals with personality disorder are assessed, treated and how effective these treatments have been. Therefore, the Peaks Unit at Rampton Hospital has set up a research programme to examine these questions. The way in which we have decided to do this is to use the information generated from patient assessments that is routinely collected by the clinical staff at the Peaks Unit. This information includes the results of clinical interviews with patients, behavioural observations of patients, and patients’ completed questionnaires. The information will be about patients’ childhood and family background, as well as their social, education, employment, criminal and mental/physical health history. This information is stored in a computer database and already used for the planning of patient treatment and service audits. By accessing this information we can answer a number of research questions.

An example might make this clearer. Suppose, for instance, that we wanted to know whether the treatment offered at the Peaks Unit meets the needs of its patients. This could be investigated by looking at the various clinical assessments that show areas in which patients have problems (e.g., anger management, impulsivity). The types of treatment provided and any changes in these problem areas during a patient’s treatment could then be examined.
It is important to point out that in order to judge whether counselling had been effective for an individual, we would be collecting information on the individual’s stress level as part of his routine clinical care in any case. In other words, this information will still be collected if you decide not to consent to the use of your information for research purposes. We are asking for your agreement that this information be added to that of other members of the patient group so that it could be used later for research purposes. The same would apply to the other types of information that are routinely collect in order to provide you with appropriate clinical care. Participation in this research project would involve you giving informed consent to our use of your clinical data for research purposes. There is nothing else you will be asked to do in relation to this research project.

**Why have I been chosen?**
In order to draw valid conclusions, ideally we would like to include all of those who are admitted to the Peaks Unit. We would therefore like to include everyone admitted. There is nothing that is unique about you in our request that you take part in this study. The results of this research will become more useful, however, if you do take part in the study.

**Do I have to take part?**
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form during your admission assessment. If you decide to take part you are still free to withdraw your consent at any time and without giving a reason. A decision to withdraw at any time will not affect standard of care you receive in the Peaks Unit.

**How will this information be stored and could I subsequently be identified when any research is published?**
It is important to reassure you that any information retained about you will be in an anonymous form so that your name will not appear on the database. Your data will only be identified by a number whose significance is known only to the researchers. The information is held on a computer and will be protected by a password that is known only to the researchers who will manage the database. The information on the database will be held for eight years. You are guaranteed that this information will not be shared with anyone who is not involved with the research. Although we would wish to publish our findings in research journals, all of the information from individuals will be gathered together as a group so that it will be impossible to identify anyone from any report that might be published.

**Risks and benefits of participation**
We are asking to use information about you that is collected by clinical staff as part of your care on the Peaks Unit. Your consent for us to do this does not require you participate in any additional procedures for the purposes of this research project, nor are there any demands on your time for being involved. Because of these facts, we anticipated no risks to you associated with your participation.
Individual patients are not expected to benefit directly from participation in this research project. We hope that results from the research project will help improve service care and planning so patients may benefit indirectly in this way from this research.

**How can I get more information?**
Professor Kevin Howells will be happy to talk to you about any aspect of the collection of these data (Tel: XXXX). Alternatively XXXXXX, the researcher who will be responsible for managing the data base (Tel. XXXX) can be contacted if you wish to discuss this further. These individuals may also be contacted by letter at the Peaks Unit, Rampton Hospital, Retford, Nottinghamshire, DN22 0PD.
APPENDIX 3

Consent Form

**Study Title:** The retention of routine clinical information on a data base for research purposes from patients admitted to the Peaks unit.

**Name of Principal Researcher:**
Professor Kevin Howells, Head of Peaks Academic and Research Unit

**Please initial box**

1. I confirm that I have read and understand the Participant Information Sheet for patients admitted to the Peaks unit, dated _______ and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that if I decide not to take part, to withdraw at any time, without giving any reason, that my medical care and legal rights will not be affected.

3. I understand that sections of any of my medical notes may be looked at by responsible individuals involved in research where it is relevant to my taking part in research. I give permission for these individuals to have access to my records.

4. I agree to take part in the above study.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of Patient</td>
<td>Date</td>
<td>Signature</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Name of Person taking consent</td>
<td>Date</td>
<td>Signature</td>
</tr>
<tr>
<td>(if different from researcher)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Researcher</td>
<td>Date</td>
<td>Signature</td>
</tr>
</tbody>
</table>

1 for patient; 1 for researcher; 1 to be kept with hospital notes