

DEVELOPING THE ACTUARIAL PREDICTION OF VIOLENT AND SEXUAL
REOFFENDING

By
PHILIP HOWARD

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

School of Psychology
College of Life and Environmental Sciences
The University of Birmingham
April 2013

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis aims to develop and improve the actuarial prediction of violent and sexual offending. It demonstrates the importance of understanding offence classification and specialisation, and the value of dynamic risk factors in actuarial risk prediction. Its findings are especially relevant to prison and probation risk assessment and management practice in England and Wales, where the National Offender Management Service (NOMS) makes extensive use of the Offender Assessment System (OASys).

Chapter 1 takes a novel, empirical approach to determining which offences should be counted as “violent” by a new nonsexual violence risk scale. Chapter 2 then develops this new scale, the OASys Violence Predictor (OVP), which combines static and dynamic risk factors, and validates it through comparison with NOMS’s existing scales. Chapter 3 then shows that OVP is also an equally good or superior predictor of nonsexual violence among offenders with a history of sexual offending. Chapter 4 shows that OVP’s dynamic risk factors have causal properties and reassessment over time improves prediction. Chapter 5 demonstrates the significance of offence specialisation by sexual offenders to risk predictor development. Chapter 6 concludes the thesis with an overview and discussion of the findings, limitations, practical implications, and future research directions.

Acknowledgements

I'm very grateful to my supervisor Louise Dixon for her support, guidance and good humour over the past six years. Louise's advice on how to write for academic publication - and also navigate university bureaucracy - has been invaluable. Our chats about rowing, triathlon and our other sporting adventures have made the PhD experience much more fun than it otherwise would have been.

Several other people have made vital contributions as both research colleagues and friends. I've greatly enjoyed working with Georgia Barnett, Ruth Mann and Helen Wakeling on chapters of this thesis, and also appreciate the support of other NOMS colleagues such as Robin Moore and Rosie Travers. Anna Strevens has provided important reality checks from an operational probation perspective, and her ability to do this while keeping me company in London's finest craft beer establishments has been especially welcome.

My family and friends outside work deserve particular thanks for their patience and encouragement over what must sometimes have seemed an endless journey. As well as my sisters, brother and mum, I'd most of all like to thank the Baileys and Marshalls in Croydon, Ruthie and Gavin in Dulwich, the Lain/Shelley family in Ilkley, Jola in Poland, Xristina in Greece, and Carol, Gina and Margaret on the other side of the Atlantic!

TABLE OF CONTENTS

Introduction	1-33
Which offences should be targeted in the prediction of ‘violent’ and ‘sexual’ offending?	1
Are violent offenders a separate group?	5
Existing approaches to the prediction of violent and sexual reoffending	8
The history of actuarial risk assessment	8
Generations of risk assessments	10
Professional judgement versus actuarial systems	11
Unstructured professional judgement	11
Structured professional judgement and conceptual actuarial systems	12
Empirical actuarial systems	12
Commentary on risk assessment typologies	13
Risk factors for violent and sexual reoffending	15
Theoretical explanations for violence	16
Empirical research on the risk factors for future violence	20

Empirical research on the risk factors for sexual reoffending	23
Age and maturity	25
Structure of the thesis	27
Aims	29
Specific aims	30
Samples	30
Statement of Authorship	34
 Chapter 1: Developing an Empirical Classification of Violent Offences.	
Chapter rationale	35
Howard, P. D., & Dixon, L. (2011). Developing an empirical classification of violent offenses for use in the prediction of recidivism in England and Wales.	37-50
<i>Journal of Aggression, Conflict and Peace Research, 3, 141-154.</i>	
Abstract	37
Introduction	37
Risk assessment in NOMS	38

The need to develop a violence risk scale for NOMS	38
The need to produce a comprehensive classification of violent offenses	39
Method	40
Overall design	40
Participants	40
Measures	40
Results	41
Determining the nonsexual violent offence content of each offence group	41
Simple associations between static and dynamic risk factors and reoffending within each offence type	44
Discussion	47
Conclusion	48
References	49

Chapter 2: The Construction and Validation of the OASys Violence Predictor (OVP)

Chapter rationale	53
Howard, P. D., & Dixon, L. (2012). The construction and validation of the OASys Violence Predictor: Advancing violence risk assessment in the English and Welsh correctional services. <i>Criminal Justice and Behavior</i> , 39, 287-307.	55-75
Abstract	55
Introduction	55
Development of a NOMS predictor of violent offending	57
Method	57
Measures	57
Procedure	59
Participants	60
Analysis	61
Results	63
Ordinal logistic regression modelling	63
Simplification of the model to a 100-point scale	63
Comparison of offender subgroups	69
Discussion	70
The development and implementation of OVP	70
Improvements in predictive validity	71
Assessment of dynamic risk factors	72
Methodological strengths and limitations	72
Conclusion	73

Chapter 3: Predicting Nonsexual Violence by Sexual Offenders: A Comparison of Four Actuarial Tools

Chapter rationale	77
Howard, P.D., Barnett, G.D., & Wakeling, H.C. (submitted). Predicting nonsexual violent reoffending by sexual offenders: A comparison of four actuarial tools.	
<i>Legal and Criminological Psychology.</i>	79-126
Title page	79
Abstract	80
Introduction	81
Method	85
Sample	85
Measures	89
Procedure	94
Results	96
Discussion	115
Limitations	119

(Note: due to reformatting during the thesis review process, pages 122-128 have been removed.)

Chapter 4: Identifying changes in the likelihood of violent recidivism: Dynamic risk factors in the OASys Violence Predictor

Howard, P. D., & Dixon, L. (2012). Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys Violence Predictor. *Law and Human Behavior*.

Advance online publication. 131-142

Abstract 131

Introduction 131

Existing evidence on the utility of measuring dynamic risk factors 132

Key considerations in the definition and measurement of dynamic risk factors 132

Existing multiwave studies of dynamic risk factors 132

Dynamic risk measurement and the Offender Assessment System 133

Method 134

Ethics 134

Participants 134

Measures 134

Procedure 135

Analysis 136

Results 137

Discussion 140

References 141

Chapter 5 – Specialization in and within sexual offending

Chapter rationale	145
Howard, P.D., Barnett, G.D., & Mann, R.E. (in press). Specialization in and within sexual offending in England and Wales. <i>Sexual Abuse: a Journal of Research and Treatment</i> .	147-173
Abstract	147
Introduction	148
Specialization in sexual offending	149
Specialization within sexual offending	151
Current study	152
Method	152
Participants	152
Measures	153
Procedure	154
Results	156
Sexual and nonsexual offending histories	156
Rates of proven nonsexual reoffending and all reoffending	160

Rates of proven sexual reoffending	161
Predictive models for proven sexual reoffending	162
Discussion	168
References	172
Chapter 6 – General discussion	175-200
Aims of the thesis, and relevant results	175
Aim One: empirical construction of a classification of offences for a predictor of violent reoffending	175
Aim Two: construction and validation of the OASys Violence Predictor (OVP)	177
Aim Three: validate OVP for offenders with a history of sexual offending	178
Aim Four: measure changes in OVP scores and subscales, and determine whether these changes are associated with changes in the likelihood of violent recidivism	180
Aim Five: understand the extent of specialisation in particular types of sexual offending and in sexual rather than general offending	182
Implications and limitations of findings and future directions for research	184
Implications of the thesis findings	184

Limitations of this research	185
Available data	185
Study length	187
New sexual offending by those with no history of this offence type	190
The validity of actuarial risk assessment	191
Future directions of research aimed at improving actuarial risk prediction	194
Complex statistical methods	194
Hazards of reoffending and offence-free time	195
Risk factors which have not yet been assessed in OASys	197
Combining actuarial assessment and clinical judgement	204
Different types of outcome measure	205
Conclusion	205
References	207-237
Ethics approval	
Letter from Dr. Robin Moore, National Offender Management Service	238
Letter from Ben Coleman, Ministry of Justice	239
Letter from University of Birmingham Science, Technology, Engineering & Mathematics Ethical Review Committee	240

LIST OF ILLUSTRATIONS

Chapter 2

Figure (2.)1	OVP score distribution of reoffenders and nonreoffenders	67
--------------	--	----

Chapter 3

Figure (3.)1	Risk Matrix 2000/V score distribution of selected offender groups, by current/prior sexual offending, age at most recent sexual offence and history of indecent image offending	103
--------------	---	-----

Chapter 5

Figure (5.)1	Combinations of offense types for those with two sexual sanctions (N = 1,988). Expected frequencies in the absence of specialization in sexual offending, frequency ranges in simulated samples without specialization, and actual frequencies.	159
--------------	---	-----

LIST OF TABLES

Table	Title	Page
Introduction		
(I.)1	Counts and date ranges of Offender Assessment System assessments, and dates of Police National Computer searches	32
Chapter 1		
(1.)1	Demographic details of samples	41
(1.)2	Indications of violent content in a range of offences	43
(1.)3	Associations between dynamic risk factors and various types of reoffending	45
(1.)4	Logistic regression models to predict proven reoffending using previous sanction counts	46
Chapter 2		
(2.)1	Logistic regression model of proven violent reoffending within 12 and 24 months	64
(2.)2	Scaling and adjustment of logistic regression model results to produce three simplified scores	65
(2.)3	Scoring of OVP from static and dynamic risk factors	66
(2.)4	Predictive validity of raw logistic regression parameters and three	67

	simplified scores	
(2.)5	Predicted probability of homicide and wounding, homicide and assault, and all violent offenses, for a range of OVP scores	67
(2.)6	Comparisons of predictive validity between OVP total score and other predictors	68
(2.)7	Sensitivity and specificity of OVP and Risk Matrix 2000/V categories	69
(2.)8	Comparisons of the predictive validity of OVP between offender subgroups	70

Chapter 3

(3.)1	Sample demographic characteristics	88
(3.)2	Four-year proven reoffending rates of offenders in the different OGRS3, OVP and RM2000/v risk categories by nonsexual violent offence and sexual offence outcome	98
(3.)3	Comparison of four-year proven nonsexual violent reoffending and sexual sexual offence outcomes by current sexual offending, indecent image offending and age at last sexual offence	101
(3.)4	Four-year proven nonsexual violent reoffending rates for selected sexual offender subgroups, by Risk Matrix 2000/v risk band and nonsexual violent offence and sexual offence outcome	105
(3.)5	Comparison of Area Under the Curve (AUC) statistics for four-year proven nonsexual violent reoffending outcomes for different violence risk assessments (using both unstandardised risk bands and total risk scores)	107

(3.)6	Comparison of four-year proven nonsexual violent reoffending outcomes for standardised banded RM2000/v, OGRS and OVP scores, using the RM2000/v total score distribution for all measures	112
(3.)7	Logistic regression models examining predictive contributions of OVP's weighted static and dynamic risk factor scores for four-year proven nonsexual violent reoffending outcomes	114

Chapter 4

(4.)1	Life table tracing violent reoffending, censoring, reassessment on OASys, and dynamic risk factor change over a 5-year follow-up	137
(4.)2	Changes in OVP risk factors between successive assessments	137
(4.)3	Initial scores and changes in score by final assessment: Reoffenders and nonreoffenders	138
(4.)4	Predictive validity of OGRS3 and initial and current OVP scores	138
(4.)5	Cox regression models: Total OVP score as a predictor of violent and homicide/wounding reoffending outcomes	138
(4.)6	Cox regression models: risk factors in OVP as predictors of violent reoffending	139
(4.)7	Cox regression models: risk factors in OVP as predictors of homicide/wounding reoffending	139
(4.)8	Acuteness of dynamic risk factors in OVP as a predictor of violent reoffending	140

Chapter 5

(5.)1	Nonsexual criminal histories, by combination of sexual offense histories	157
(5.)2	History of each sexual offense and proven reoffending rates for nonsexual and all offenses	160
(5.)3	History of each sexual offense and proven reoffending rates for four sexual offense groups and five aggregate sexual offense groups	162
(5.)4	Cox regression models of four simple sexual reoffending groups	164
(5.)5	Cox regression models of five aggregate sexual reoffending groups	165
(5.)6	Concordance Indices for nine predictive sexual reoffending models, applied to similar sexual reoffending outcomes	167

INTRODUCTION

This thesis deals with a familiar topic in forensic psychology and criminology research: the identification of offenders who are likely to offend again. Research efforts, especially in the academic rather than government scheme, are particularly focused on violent and/or sexual reoffending. These offences are deemed to be most harmful, and therefore are the focus of forensic psychological practice. Developments in risk assessment instruments which identify potential violent and sexual recidivists can therefore bring considerable public benefit, through improved targeting of public protection and rehabilitative efforts. This Introduction underpins five Chapters which directly develop risk assessment instruments in this way and/or aid future developments by improving understanding of how static and dynamic risk factors are related to offending. As such, the Introduction presents an overview of the existing theories and empirical evidence on risk factors for violent and sexual recidivism. It also presents some material on whether offenders specialise in these offences, and on existing risk assessment tools, though without duplicating material in the main Chapters.

Which offences should be targeted in the prediction of 'violent' and 'sexual' reoffending?

While violent and sexual reoffending have been extensively studied in criminological and psychological research, the boundaries of what constitutes “violent” offending - including whether it should include sexual offences - have still not been clearly established.

A broad view of aggressive behaviour could incorporate many acts. Parrott and Giancola (2007) raised a criterion problem in much research on aggression, which has “inconsistently

differentiated between *aggressive behaviour* and the related *emotional* and *attitudinal/cognitive* constructs of anger and hostility, respectively” (p.281). They noted that as far back as the work of Harre and Lamb (1983), “over 200 different definitions of aggression [had] been advanced” (p.282). Clearly, for purposes of crime research and reduction, the primary focus must be placed on behaviour. A distinction has long been drawn between instrumental aggression, where aggression is directed to achieve goals such as material gain through robbery, and unplanned behaviour variously described as angry, annoyance-motivated or hostile aggression (Blackburn, 1993). This was criticised by Bushman and Anderson (2001), as many aggressive acts include a mixture of both types of behaviour. In its place, Parrott and Giancola (2007) proposed a taxonomy combining two classifications, with acts being either active or passive and either direct or indirect. Subtypes of each of the four groups (e.g. active direct) are physical, verbal, damage to property, theft and (for active acts only) postural. In any case, the perpetrator must intend harm to the specific victim, although this harm can take a range of forms such as inconvenience (e.g. through damaging or stealing a means of transport), loss of social or economic standing (e.g. spreading harmful rumours, deliberate withdrawal of support in a work situation) or psychological discomfort (e.g. insults) as well as causing physical injury or pain.

Whatever typology of aggressive behaviour is accepted, not all aggressive acts are criminal. For example, most of Parrott and Giancola's passive acts are wholly legal (e.g. insults which do not encompass physical threat) or only civil offences (e.g. maltreatment by an employer; most acts of slander). The aggression also has to be non-consensual to be criminal: harm caused as a byproduct of benevolent surgery, contact sports or consenting sado-masochistic acts is unlikely to be considered illegal or to be prosecuted where it is illegal.

Existing research on violent recidivism has tended to pay little attention to this broad range of aggressive behaviours, and therefore lacks rigour in one important respect: defining the outcome of interest. Examination of a meta-analysis by Campbell, French, and Gendreau (2007) is illuminating. Of their initial 88 manuscripts suitable for meta-analysis, 59 studied violent reoffending as well as, or instead of, institutional misconduct. Of these, I have scrutinised 40 manuscripts which have been published in peer-reviewed journals or are available online. Of these 40, six contain no definition of violence whatsoever, while three others state only an interest in 'offences against persons' and two are actually related to 'serious'/'less serious' offences or restricted to domestic abuse. Four are explicitly linked to the definitions in risk assessment tool manuals. After some grouping of studies due to shared authorship and common methods, 18 definitions can be studied. Findings included that: all definitions included some nonsexual assaults, but there was variation on whether noninjurious assault should be included; sexual offences could be always or never included, or included when it involved any victim contact or only when the offence was rape; all robbery was included in about half of the definitions, with many others including armed robberies and/or robberies with force; offences such as threats to kill, other threatened violence and harassment were not specifically mentioned in many sources, and otherwise were sometimes included, sometimes excluded and sometimes restricted to 'serious threats'; half of the studies did not mention weapon-related offences at all, while others dealt variously with weapon possession and the active pointing or use of weapons, or failed to clarify which kinds of weapon-related offences they were discussing; and, criminal damage offences were rarely considered at all, with some authors mentioning arson but none including non-arson criminal damage, even where there was a threat to the person.

This wide variation in definitions of violence is potentially relevant to the success of efforts to predict recidivism because of offence specialization. If offenders have tendencies to commit certain types of offence which last long enough that past offending and/or measured risk factors are a guide to their future offence preferences, then risk prediction tools which focus on particular types of recidivism will be more successful if they are focused on groups of offences which 'sit together'. To take two major elements of the above definitions, if the set of offenders who commit nonsexual assaults overlaps by much more than chance with the set of offenders who commit sexual offences, and the risk factors associated with both types of offending are similar, then attempts to predict violent and sexual reoffending with the same prediction scale are likely to be successful. If not, then – given that nonsexual assault is far more frequent than sexual offending (see, for example, the frequencies in Chapters 1, 3 and 5) in terms of both criminal history and recidivism – the prediction of nonsexual assault will overshadow, in the arithmetical sense, the prediction of sexual offending, and the scale is therefore unlikely to predict sexual reoffending well.

As such, this thesis takes a strong interest in offence specialization. Chapter 1 follows up on the variation in violence definitions identified above, and is dedicated to establishing which offence types can successfully be added to a 'core' group of nonsexual assault offences, as a precursor to the production of the OASys Violence Predictor in Chapter 2. It establishes, among other findings, that sexual offences should indeed not be the target of the same predictor as nonsexual violent offences. This concords with an important recent meta-analysis of the prediction of sexual, violent (including sexual) and any recidivism among sexual offender. In this study, Hanson and Morton-Bourgon (2009) confirmed that predictive validity

is maximised when risk predictors are used for the outcome for which they are designed.

Among their findings, measures designed for sexual recidivism only were better predictors of sexual recidivism than measures designed for both sexual and nonsexual violent recidivism.

Chapter 5 returns to sexual offences to determine whether attempts to predict all sexual reoffending with a single measure are likely to succeed, or whether instead the (typically) less harmful noncontact offences should be 'discarded' in order to safeguard prediction of the (typically) more harmful contact offences.

Are violent offenders a separate group?

Many offenders commit both violent and non-violent offences. Contemporary research therefore seeks to distinguish whether specialist “violent offenders” exist as a significant and potentially identifiable group. The answer is not yet clear: “the primary focus of research on this topic has been discovering whether any more specialization exists than would be expected by chance alone. Research to date is disappointingly unclear on this point” (Osgood & Schreck, 2007, p. 274). This topic is clearly relevant to the prediction of violent reoffending: if specialisation does not exist, then only predictors of general reoffending are required.

Farrington (1999) drew on the Cambridge cohort study to present the view that violent offending is best understood in the context of general offending: “Offending is predominantly versatile rather than specialized, particularly at younger ages... violent offenses appear to occur almost at random in criminal careers” (p. 155-156). Against this, Brame, Bushway, Paternoster, and Thornberry (2005) found that males in the Rochester Youth Development

Study exhibited a prior behaviour effect: those with a self-reported violent offence in one time period were about twice as likely than others to self-report a violent offence in the following time period.

Soothill, Francis, Ackerley, and Humphreys (2007) found that young adult males recorded on the English and Welsh Offenders Index fell into 16 clusters with similar offending patterns. The nine single-offence clusters included a large criminal damage cluster and a small resisting arrest cluster, and the three dual-offence clusters included a large violence/criminal damage cluster. However, there were also three medium to large-sized multi-offence clusters where violence and criminal damage featured alongside other types of offending. Among the five clusters for female offenders were a specialist violence cluster and a 'versatile' cluster which included violence and criminal damage. In a similarly large Danish sample (Brennan, Mednick, & John, 1989), there was some specialisation in violence among those with more than three arrests. Not all specialization research takes violent or sexual reoffending as a starting point or even, like Soothill et al. and Brennan et al., takes an equal interest in all offences. For example, a limited amount of research has typologised burglars, most recently creating four categories with varying criminal history extent and versatility (Hahn Fox & Farrington, 2012).

Mazerolle, Brame, Paternoster, Piquero, and Dean (2000) sought to test the distinction made by Moffitt (1994) between “life-course persistent” and “adolescent limited” offenders. This distinction proposes that the former should demonstrate long, serious, versatile criminal careers, while the latter (who include most female offenders) should be involved in more specialised careers which start later, are shorter and unlikely to involve violence. Using data

from the Philadelphia Birth Cohort on offending to age 26, Mazerolle et al. confirmed that offending versatility is greater among those with earlier onset for both men and women. This result suggested that violent offending is often part of a wider pattern of offending by members of the life-course persistent group. However, Sullivan, McGloin, Pratt, and Piquero (2006) found, for a sample of serious felons, that the apparent degree of offending specialization is sensitive to the choice of measure used. They focused on monthly self-reported offending histories, instead of the year-by-year or longer-term analyses usually conducted. They report that “studies spanning multiple years reveal that most offenders – and especially chronic offenders – are not terribly picky when it comes to the type of criminal acts they are willing to commit” (Sullivan et al., 2006, p. 220), but, while confirming the frequency-diversity link, “the offenders in this sample evidenced a considerable amount of short-term specialization” (p. 221).

Osgood and Schreck (2007) offered a thorough description and criticism of the methodologies used in earlier studies to assess the degree of specialization in a sample's offending history. They developed a method which isolates the degree of specialization in violent offending within a multilevel regression model based on item response theory, i.e. controlling for the varying base rates of each offence type and offenders' varying propensity to commit any type of offence. Applying this model to three juvenile datasets found “no doubt that that individual differences in specialization in violence are greater than can be accounted for by chance alone” (Osgood & Schreck, 2007, p. 292). These individual differences in specialization comprised “anywhere from about half to all of the magnitude of the variance in overall offending”. Comparing individuals' residuals across successive study waves showed that

specialization levels were fairly stable over time¹. McGloin, Sullivan, and Piquero (2009) revisited the evidence and used latent transition analysis to attempt to determine whether offenders moved between offence classes over time. They found tentative evidence that some offenders might specialise in the short term (within one-year periods) before moving on to other offence classes, but were unable to reach a strong conclusion and urged further research including methodological development.

As such, the balance of existing evidence suggests that some offenders convicted of violent offences specialise in such offending, and therefore there should be merit in maintaining separate predictors of violent reoffending. Still, the possibility that general predictors could work well for violent and sexual reoffending is not ignored, with the predictive validity of the static general reoffending scale OGRS3 being checked for these outcomes in Chapters 1 and 5 respectively. Material on specialisation in and within sexual offending is included in the introduction to Chapter 5.

Existing approaches to the prediction of violent and sexual reoffending

The history of actuarial risk assessment

This paragraph is indebted to the historical exposition presented by Harcourt (2007).

Statistical approaches to the prediction of reoffending were first conceptualised in the 19th

¹ Potential explanatory variables which may affect overall offending propensity and/or violence specialization can be added to the model. Osgood and Schreck found several significant relationships with violence specialization, although with some differences between different datasets; these parameters differed from those for offending propensity, suggesting that there is not a straightforward correspondence between overall offending propensity and specialization in violence.

century but were first applied to generate risk scores for the Illinois parole system from 1935 onwards (Burgess, 1936). This system scored offenders on the presence or absence of 21 dichotomous factors which were empirically associated with success on parole. Following, research in various papers by Glueck and Glueck used a narrower range of risk factors and replaced the simple “Burgess method” by giving different weights to different risk factors (“Glueck method”). This more complex approach held the promise of improving prediction, but did not achieve universal coverage – as more parole risk predictors and, later in the century, pre-sentence and custody classification predictors were developed, a mix of Burgess and Glueck scoring emerged. Indeed, in comparisons of methods (e.g. Silver, Smith, & Banks, 2000), the Burgess method often predicted as well as the Glueck method or more complex methods, which now include scores based on logistic regression and simple and iterative classification trees. While the US federal government uses a seven-item Burgess method, the Salient Factor Scale (SFS), most US jurisdictions using structured risk predictors now prefer more complex methods such as the Level of Service Inventory – Revised (LSI-R; Andrews & Bonta, 1995).

In the UK, probation services were only required to use structured risk assessment methods in 1992; a near-duopoly developed by 1998, with the imported LSI-R, and the homegrown Assessment and Case Evaluation (ACE) in use in 20 and 25 of the (then) 54 local probation services respectively (Raynor, Kynch, Roberts, & Merrington, 2000). Both systems were driven out of use early in the next decade by the Offender Assessment System (OASys; Home Office, 2002), which became mandatory in all probation areas by 2003.

A range of approaches are now in use. As the purpose of assessments has moved away from

'pure' prediction to identification of treatment targets, an emphasis has been placed on dynamic risk factors (i.e. those with the potential to change, especially as a result of the efforts of correctional officers). There are therefore now several competing styles of risk assessment. These styles can be classified in at least two ways.

Generations of risk assessments

Risk assessment methods can be seen as having changed over time, and are accordingly classified into generations (Bonta, 2008; Andrews & Bonta, 2003). Each generation attempts to remedy the shortcomings of its predecessor.

First generation risk assessments are purely based on clinical judgement, with no role for actuarial methods (scoring). The assessor decides which factors should be assessed and informally weighs these factors in classifying the offender.

Second generation risk assessments arose because pure clinical assessments were considered to lack reliability and predictive validity. These are predominately based on static indicators – typically age, gender and criminal history – which have strong predictive validity. Tools for general reoffending include SFS and the contemporary English and Welsh tool Offender Group Reconviction Scale version 3 (OGRS3; Howard, Francis, Soothill, & Humphreys, 2009). Also used in England and Wales is Risk Matrix 2000 (Thornton et al., 2003), which has scales for both sexual and violent reoffending.

Third generation risk assessments, such as LSI-R and the Historical Clinical Risk - 20 (HCR-

20; Webster, Douglas, Eaves, & Hart, 1997), combine static and dynamic risk factors. Unlike second generation assessments, they therefore allow the offender's score to change as their socioeconomic, environmental, interpersonal, cognitive and substance-related problems change.

Fourth generation risk assessments, such as OASys and the Level of Service Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2006), are the most recent development. They build on the considerable strengths of third generation assessments by adding a mechanism for sentence / case management plans to be based on the identified risk factors and responsivity issues.

Professional judgement versus actuarial systems

Despite the long history of actuarial prediction, it has not replaced the judgement of probation officers and other corrections staff. The majority of assessment tools still require some degree of clinical decision making. The division between 'clinical' and 'actuarial' tools is therefore blurred. Hanson and Morton-Bourgon (2007) made a distinction between 'professional judgement', where scores may or may not be included but the final risk classification is made by the clinician – and actuarial systems, where the final classification is based on scores. These two groups become four, as professional judgement may be structured or unstructured and actuarial systems may follow a 'conceptual' scheme or be entirely empirical. Empirical actuarial scores appear to have the greatest predictive validity. (In an update of this meta-analysis (Hanson & Morton-Bourgon, 2009), conceptual actuarial tools became 'mechanical' and a further analysis suggested that allowing professional override of empirical actuarial

scores is counterproductive.)

Unstructured professional judgement

Unstructured professional judgement is, by its nature, not a system at all. The assessor is free to draw their own inferences from whatever qualitative (e.g. interview, case conference) and quantitative (e.g. psychometric test) data they choose. While the prevalence of unstructured judgement in assessment of criminals has declined in recent decades, due in part to an accumulation of evidence on its poor predictive validity (see later), it was still used in over one-quarter of US jurisdictions in a 2001 survey (Harris, 2006). It is equivalent to the first generation described by Bonta (2008).

Structured professional judgement and conceptual actuarial systems

In structured professional judgement (SPJ) systems, the assessor is required to consider a list of items, which generally require scored responses. The final risk classification, however, is their decision. A feature of these systems is that a conceptual framework operates. That is, the items are chosen for their clinical significance and theoretical underpinning at least as much as for their predictive validity. They are usually grouped into subscales which capture particular facets of risk. Conceptual actuarial systems choose and organise items in a similar way, but differ from SPJ in that the final risk classification is produced mechanically from the item scores. As Hanson and Morton-Bourgon (2007) noted, an SPJ system will become a conceptual actuarial system when researchers “omit the professional judgement and simply add the items from the checklist” (p. 3).

Empirical actuarial systems

In empirical actuarial systems, the items included are chosen because of the strength of their relationship with the outcome. It is not necessary for a conceptual framework to explain why the items have been included – for example, by explaining the processes by which the presence of the risk factors leads to offending behaviour – nor organise the risk factors into clinically useful families. While all of Bonta's (2008) second generation tools are empirical actuarial, some tools which include dynamic risk factors, such as the Violent Risk Assessment Guide (VRAG; Quinsey et al., 2006) are also empirical actuarial.

Commentary on risk assessment typologies

The generation system is interesting, but the historical succession it implies seems factually dubious. Some of the earliest prediction systems described above included dynamic risk factors. These third-generation systems could therefore be said to have arisen at the same time as or even predate the second-generation systems. For example, research by Hart (1923) included factors such as “extent of occupation 'regular'... using cigarettes... claims to attend church regularly” (Harcourt, 2007, p.274-5). The 'generation' schema seems to say more about how risk assessment theory and practice changed over time rather than describing individual assessment tools. The third-generation group is extremely broad, and covers almost all systems in use for rehabilitative work (as opposed to initial screening work). It is not clear what such a broad schema achieves. The fourth-generation OASys, now used to assess and manage risk in probation and prison in England and Wales, essentially combines a mixture of

second- and third-generation subsystems with a sentence planning structure.

The professional judgement / actuarial split says much more about what contemporary assessment systems do, and thus distinguishes between various third- or fourth-generation tools. Douglas and Skeem (2005) made an interesting point about the content of the tools which fall into different categories: “although actuarial guides could, in theory, include causal dynamic risk factors, extant risk factors heavily weight static variables, nearly to the exclusion of dynamic variables... the SPJ approach tends to include greater emphasis on dynamic risk factors” (pg. 352).

The dividing line between SPJ and conceptual actuarial systems is, however, less clear than it might first appear. For example, while the Psychopathy Checklist - Revised may not draw its own dividing line between psychopaths and non-psychopaths, “categorical diagnoses of psychopathy may be useful or required for some research or clinical applications” (Hare, 2004, p.30) and cut-off scores are therefore promoted to encourage consistency in risk assessment and treatment decisionmaking.

As the principal data source for much of this thesis, the Offender Assessment System's design – described in Chapter 2's Methods section - is of considerable importance. It contains both static and dynamic elements, and some items are scored through clinical judgement (with varying degrees of structured guidance given in the OASys manual). Its division into sections was conceptual, while the Glueck weights given to the different sections in the original OASys score give this actuarial scale both conceptual and empirical elements, given that the weights were determined through informed but unsystematic synthesis of the evidence

available at the time. (This process is vaguely referred to in the OASys manual but, having joined the OASys team in what was then part of the Home Office soon after the initial rollout of OASys, I have also been told the 'secret history' of OASys by more experienced colleagues.) The OASys Violence Predictor (OVP), whose design is also related in Chapter 2, looks like an empirical actuarial scale, but there was a conceptual aspect to its design, as understanding of the risk factors for violent offending was drawn upon to create a longlist of OASys items which were then used in the empirical process of creating the new Glueck weights. The Glueck/Burgess debate is also considered in Chapter 2, as the effects upon predictive validity of smoothing Glueck weights slightly for userfriendliness and moving to a more Burgess-like system of equal weights for each risk factor are both evaluated.

Chapter 4 is also relevant to risk assessment typology, with respect to the issue of dynamic risk factors. It addresses Douglas and Skeem's point about dynamic content in actuarial systems, by investigating whether the supposedly dynamic risk factors in the OASys Violence Predictor (see Chapters 1 and 2) do demonstrate causal dynamic properties. It includes a review of several similar, but far smaller, studies of sexual and violence risk prediction.

Risk factors for violent and sexual offending

In this section, a number of theoretical explanations for violent offending are summarised, followed by a description of empirical evidence around risk factors for violence. While many of these theories and factors have at least some relevance to sexual violence as well as nonsexual violence, a further subsection summarises recent empirical evidence on risk factors for sexual offending. Finally, the topic of age – a risk factor which tends to transcend

theoretical schools - is discussed.

Many of these empirically supported risk factors are components of the models of violent and sexual offending developed in the chapters of this thesis, while others lacked support in these data analyses and others could not be considered due to the limitations of the available data sources. These results are considered further in the concluding Discussion. It would be possible to develop and validate risk tools without knowledge of the existing empirical evidence and underlying theories. However, an analyst taking this approach would lack appreciation of the relationship between predictive scores and treatment models, could fail to acknowledge any idiosyncrasies in their results, and would not be aware of the potential to develop their work further by gathering data on a wider range of risk factors.

Theoretical explanations for violence

A wide range of theories have been developed in order to explain why violent behaviour occurs and identify likely risk factors. They are founded on the different approaches followed in several disciplines of psychology and other social and life sciences. A selection of these theories and supporting evidence are now summarised briefly, as the main focus of this thesis is upon improving the empirical rigour of risk assessment instruments.

Biological theories, such as the evolutionary model of Wilson (1978), stress similarities between humans and animals. As Blackburn (1993) explained, some violent behaviours and the associated neurobiological processes, are shared across species and are observed across all human cultures – these include the 'fight or flight' reaction and the role of the limbic system in

integrating emotional arousal and expression. Neurological research has developed rapidly in recent years, and could potentially play a more prominent role in the assessment of individual offenders in the future, if it becomes both affordable and reliable (Beech, 2008). Evidence on specific systems such as the roles of the neurotransmitter norepinephrine (Haden & Scarpa, 2007) and testosterone (Archer, 2005) is already accumulating. A variant on the biological theories proposes that the nutritional requirements of the brain are not always met adequately in either everyday life nor correctional institutions, resulting in antisocial behaviour including violence. A double-blind study providing nutritional supplements or placebos to British young adult prisoners found that supplementation led to a 35% reduction in disciplinary offences (Gesch, Hammond, Hampson, Eves, & Crowder, 2002). Williams (2012) summarises research on another physiological risk factor: traumatic brain injury (TBI). Brain injury leads to “loss of memory, loss of concentration, decreased awareness of one's own and others' emotional state, poor impulse control and, particularly, poor social judgement” (p. 11), and can therefore disinhibit offending behaviour, with right front injuries especially linked to impulsive aggression. TBI can limit ability to benefit from forensic rehabilitation unless adaptations are made to compensate for these difficulties. Worldwide studies find general population prevalence of TBI below 10%, compared with 50-80% among offender populations; for example, 60% of male prisoners in HMP Exeter reported a “head injury”. Cohort studies in Finland and Sweden have associated brain injury with perpetration of violence, with the Swedish study including a control for genetic, social and economic risk factors: the TBI population had heightened violence risk when compared with their uninjured siblings.

To understand when aggressive behaviour will occur in social situations, researchers have designed experiments to test individuals' responses to particular scenarios. In Archer (2007),

prisoners were presented with scenarios which described a competitor's size, allies and reputation. They were most likely to respond violently to insults from competitors who had similar Resource Holding Power – ability to engage in physical conflict - to themselves, and thus were threatening but might be defeated. Men were generally more likely to respond physically, and placed more importance upon reputation. Miller and Maner (2008) examined responses to imagined infidelity. Males were more likely to experience anger and respond violently, especially towards their same-sex competitor, while women were more likely to be violent (if they were violent at all) towards their unfaithful partner. The biological theories and these experiments suggest that exposure to conflict situations may be a risk factor, while nutrition and brain injury could have validity as separate risk factors.

In social learning theory (Bandura, 1977), the process of modelling (social learning) occurs when the subject pays attention to a model carrying out a particular behaviour, retains details of what the model has done, is physically capable of reproducing it, and has the motivation and opportunity to do so. This process is not restricted to violence, but to a wide range of prosocial acts also. Relevant experiments have often featured young children observing and copying adult behaviour. A related field of research is around media violence, where it has been proposed that violence in media such as television and computer games might encourage young people to engage in violence in real life. Individuals who are more susceptible (i.e., have problems on other risk factors) are held to gain violent cognitive scripts, become physiologically aroused and have less control on natural tendencies to imitative behaviour (Anderson & Dill, 2000). Sceptics have criticised the methodological quality of empirical studies in this area and found that, at a community level, levels of overall and youth violence do not correlate with consumption of media violence (Olson, 2004). In general, social

learning theory suggests that peer and parenting factors could be related to violence.

The frustration-aggression hypothesis was summarised and revised by Berkowitz (1993), covering its long history since its inspiration by Freud. It states that each aggressive act is the result of prior frustration. However, frustration does not always lead to aggression, whether direct physical aggression or more indirect forms such as verbal aggression or fantasy.

Berkovitz suggested that anger was a crucial intermediate variable. As well as physical provocation and insults, economic problems such as unemployment can lead to frustration and therefore aggression. The final version of this model, relabelled “cognitive reassociation”, suggests that negative affect leads to bodily arousal, which firstly pushes the individual towards a violent reaction and secondly primes them to interpret subsequent events in a manner which is consistent with their current cognitive disposition towards violence – that is, there is feedback between instinctive and cognitive responses. Risk factors associated with this model should include perspective taking ability and hostile attribution. When testing whether these behavioural scripts lead individuals towards pro-aggression interpretations of potentially violent situation, normative beliefs about violence were found to be related to variance in aggression over and above trait anger (Gilbert, Daffern, Talevski, & Daffern, 2013). Attitudes towards violence may therefore be a further cognitive risk factor.

In theories of anger (Davey, Day, & Howells, 2005), violent offenders can be undercontrolled - impulsive and lacking inhibition, therefore committing readily to violent courses of action – or overcontrolled, with strong inhibition but potential build-up of negative emotions which can lead to extreme, explosive violence. Some of these offenders will deny their experience of anger, while others will be inhibited, recognising their anger but ruminating upon and

rehearsing their grievances rather than expressing them. Different treatment strategies are likely to be successful for each group, though firm conclusions have not yet been reached.

Empirical research on the risk factors for future violence

The theories outlined above outline a range of risk factors. Some authors have sought to assess the relative relevance of these risk factors by assessing the strength of their empirical support. These can be based on meta analysis: the combination of results from large numbers of previous studies, applying certain statistical formulae, in order to produce robust estimates of the strength of the effect of predictors or risk factors which are found in many or all of these previous studies.

Andrews and Bonta (2003) drew upon an ongoing meta analytic project at the University of New Brunswick and Carleton University in which “approximately 1,000 studies had been listed, 700 studies located, and 373 studies subjected to content analysis and meta analysis [yielding] more than 1,770 Pearson correlation coefficients” (p. 75). The results of this project led them to identify eight risk factors for general reoffending. The “Big Four” are antisocial attitudes, antisocial associates, a history of antisocial behaviour and an antisocial personality pattern. The other four, which are less strongly correlated with reoffending, are problematic family/marital circumstances, problematic school/work circumstances, problematic leisure circumstances and substance abuse. Risk factors associated with social class or parental achievement, and with personal distress, are far less strongly correlated with recidivism.

Douglas and Skeem (2005) presented a review of promising dynamic risk factors – those where empirical evidence supports their link with violence and their ability to change over time. (They also said more about the evidence available at the time on whether changes really occur and the methodological issues associated with such study – this topic is considered thoroughly in the introduction to Chapter Three.) These risk factors were:

Impulsiveness. This is often targeted by treatment programmes for violence offenders, but there is “limited empirical support” for its link with violence (Douglas & Skeem, 2005, p.359). Conceptually, it causes inability to stay calm under pressure and can therefore lead to inadequate self-control among those with strong temper. The work of Barratt (1994) is important, and subsequent research with his Barratt Impulsiveness Scale (BIS-11) suggests that proneness to impulsive behaviour is dynamic.

Negative affectivity. Anger is discussed above, though Douglas & Skeem note meta-analytic findings suggesting that it is malleable to treatment. Some evidence suggests that anxiety, depressive symptoms and neurosis predict violence; plentiful evidence shows that mood and affect are liable to change in both the long- and short-term. Conceptually, negative mood may lead to negative cognitions about self and others, irritability and impulsivity. It can also affect and be affected by other socioeconomic, relationship, substance misuse, cognitive and mental health problems.

Psychosis. While psychotic symptoms have consistently been shown to be dynamic, evidence on the link between psychosis and violence is decidedly mixed. The most promising

link involves those symptoms which combine loss of self-control (e.g. hearing voices which give commands) and feelings of being threatened (e.g. persecution beliefs). The frustration and anxiety associated with psychotic problems may aggravate the effect of other risk factors.

Antisocial attitudes These are strongly related to general recidivism (Andrews & Bonta, 2003), and some findings also refer to violent recidivism. Attitudes conducive to violent behaviour tend to facilitate such behaviour, and can change over time as a result of persuasion or formal treatment programmes.

Substance misuse A great deal of evidence shows that violence is more likely among those who misuse alcohol and perhaps some illegal drugs. Episodes of substance use are necessarily dynamic, and patterns of use also ebb and flow. Epidemiological links between substance misuse and violence are strong, and moreover a causal link is demonstrated by some studies which look at the exact timing of drinking and violent acts. Substance use is likely to act as a disinhibiting agent, and also worsen other risk factors. Other explanations posited by Douglas and Skeem (which seem unlikely to explain most substance-related violent episodes) include aggression being required in drugs transactions, frustrations associated with failing to obtain/use substances, and substance use being only correlated with violence through the operation of a third factor (e.g. neighbourhood, psychiatric disorder).

Interpersonal relationships Offenders are more often violent towards friends and relatives, and the quality of relationships is associated with domestic violence. Several studies (measuring the nature of relationships in various ways) confirm that relationships can also act as a protective factor, when others provide material and/or emotional support which can help

the offender to cope with their problems. Understanding the complex nature of social networks is crucial in order to make use of this risk/protective factor.

Treatment alliance and adherence. Co-operation with mental health treatment – both in terms of taking medication and meeting with mental health professionals - correlates with lower levels of future violence. Treatment and medication may directly protect against violent behaviour or, as with other risk factors, indicate higher levels of other, causal, risk factors.

Evidence on the role of neuropsychological executive functioning is rapidly accumulating. The latest meta-analysis (Ogilvie, Stewart, Chan, & Shum, 2011) compared the associations of a range of executive functioning disorders with a range of measures of antisocial behaviour. Moderate differences in executive functioning were found between criminals and noncriminal controls. Moderate effect sizes were also found for the juvenile conditions of conduct disorder and oppositional defiant disorder, which can be precursors to persistent antisocial behaviour in adulthood.

Empirical research on the risk factors for sexual reoffending

Mann, Hanson, and Thornton (2010) reviewed studies which had identified risk factors for sexual reoffending, combining two previous meta-analyses with the results of two newer, large studies. They ranked factors as empirically supported, promising, unsupported overall but with interesting exceptions, worth exploring or having little or no relationship to sexual recidivism. Empirically supported factors were sexual preoccupation, deviant sexual interest (of which the most promising was penile plethysmograph evidence of sexual interest in

children), offence-supportive attitudes (although the consistent assessment of this risk factor is described as problematic), emotional congruence with children, lack of emotionally intimate relationships with adults, lifestyle impulsivity, general self-regulation problems, poor cognitive problem solving, resistance to rules and supervision, grievance / hostility, and negative social influences. Promising factors were hostility towards women, Machiavellianism, callousness / lack of concern for others, and dysfunctional coping. Factors which were unsupported overall but with interesting exceptions include denial, view of self as inadequate, major mental illness and loneliness. Factors unrelated to sexual recidivism were depression, poor social skills, poor victim empathy and motivation for treatment at intake. Factors worth exploring – on the basis that limited studies had presented promising findings - were adversarial sexual orientation, fragile narcissism and sexual entitlement. Mann et al. emphasised that even for the empirically supported risk factors, “considerably more work is required, however, to establish their causal connections with recidivism”(pg 208), including studies where changes in the risk factors are deliberately induced. The partial but not complete overlap of their list with risk factors for general recidivism is acknowledged. Reliable measurement of several risk factors and the establishment of thresholds for significant change are identified as potential problems. In a more recent meta-analysis (Helmus, Hanson, Babchishin, & Mann, 2013) attitudes supportive of sexual offending were found to be significantly associated with sexual recidivism, probably more so for child molesters than rapists. Nevertheless, Helmus et al. cautioned that “further clarification and understanding of offense supportive attitudes is necessary, including its relationship to other constructs, such as denial/minimization, general procriminal attitudes, sexual deviance, and other offense-related attitudes (e.g., hostility toward women)” (p. 15-16). As a further example of how sexual offenders cannot be approached as a single group with uniform risk

factors, the meta-analysis by Babchishin, Hanson, and Herman (2011) found that online sexual offenders had higher levels of sexual deviance, but also more victim empathy and lower levels of impression management, than offline sexual offenders.

As a side note, Chapters 3 and 5 do not refer to “child molesters” and “rapists”. I consider these terms unnecessarily emotive, and also difficult to apply to real English and Welsh data in that they do not cover noncontact offending while appearing to double-classify the crime of raping a child. Instead, to proceed carefully towards an understanding of specialisation within sexual offending (Chapter 5), relevant offences are classified comprehensively and using neutral language. These comprise contact sexual offences against children and against adults, indecent images offences (i.e., sexual images of children, as described in statute) and paraphilias (other noncontact offences, most frequently indecent exposure or voyeurism).

Age and maturity

Age is a highly-weighted risk factor for recidivism in many empirical actuarial risk assessment tools. Examples include OGRS for general reoffending, VRAG and Risk Matrix 2000/V for violent reoffending, and Risk Matrix 2000/S and Static-99 (Hanson & Thornton, 2000) for sexual reoffending. This finding has had mixed backing in a range of subsequent studies (see Craig, 2011, for a review related principally to sexual offending). In my view, many of these studies have either not properly controlled for other risk factors (R. Karl Hanson, personal communication, 8 Sep 2012), involved too-small sample sizes and/or used convoluted techniques to control for age which ultimately do not improve prediction (e.g., Barbaree, Langton, Blanchard, & Cantor, 2009, which partials out the associations between

age and sexual deviance and antisocial behaviour and then adds an age term back).

Given that an offender's age is wholly outside their control, ethical considerations make it especially important to demonstrate convincingly why age should be associated with recidivism, and why alternate factors for which the offender could have some degree of personal responsibility cannot be used instead with equal predictive validity.

Prior et al. (2011) reviewed the literature on maturity and offending, as it relates to offenders aged 18 and over. They suggest that emotional and social development continues during early adulthood, whereas physical and intellectual development tends to have been completed by age 18. From a risk factors perspective, continuing changes in the risk factors associated with incomplete individual development are accompanied by changes in peer relations and socialization processes. (See my note above that exposure to conflict situations may be a risk factor: intimate and peer relationship formulation and dissolution, and night-time socialization, are more frequent among younger adults.) Neurological research (e.g., Edwards, 2009) provides evidence that changes in executive functioning do not cease until many individuals are into their early- to mid-twenties, and several studies of psychosocial capability (e.g., Modecki, 2008) agree that these include temperance processes such as evaluating the consequences of actions, and limiting impulsivity and risk-taking. (These can be considered similar to the 'self control' of the general theory of crime: see Gottfredson & Hirschi, 1990.) Moffitt (2006) suggested that adolescent-limited offenders (those whose offending does not continue into later adulthood) will cease to offend as the social, calendar and legal restrictions placed upon them are gradually lifted as their calendar age increases. Wikstrom and Svensson (2010) reported interactions between individuals' personal risk factors and the extent to which

they are exposed to a criminogenic environment in determining their propensity to offend.

For risk factors associated with maturity to be used in risk prediction, they must be assessed.

Prior et al. (2011) consider that the Offender Assessment System includes some items related to temperance and the other maturity-related risk domains, perspective and responsibility.

They conclude that OASys and the youth justice tool Asset offers a “partial means of assessing the maturity of offenders” (p. 29) but that “there would remain potential areas of inconsistency in [their] application” (p. 30). In the present author's judgement, this partial coverage of maturity issues in OASys creates a gap which may have to be filled by the use of calendar age as a risk factor, if the available individualised measures are inadequate to fully capture the extent to which younger adults have more criminogenic personal risk factors and are exposed to more criminogenic environmental influences than older adults. Paper 2 includes age as part of the OASys Violence Predictor, despite the presence of a broad range of static and dynamic risk factors; Paper 5 includes exploratory modelling of the risks of different types of sexual reoffending, including age together with a range of static risk factors, though dynamic risk factors were not available.

Structure of the thesis

This thesis principally investigates two themes related to the prediction of violent and sexual reoffending. Firstly, the role of dynamic risk factors in nonsexual violence risk prediction is studied. Secondly, research which promotes the understanding of offence specialization as a necessary condition for successful empirical actuarial prediction is provided. Given the intertwining nature of these themes, the five papers which follow have been organised using

the following structure.

Part I is concerned with the construction and validation of a new violence risk predictor for use in NOMS, with offenders assessed using OASys. Part II looks particularly at whether this OASys Violence Predictor includes causal dynamic risk factors. As Part I showed that sexual offences should not be included in OVP, Part III investigates specialization within sexual offending, as well as considering whether sexual offenders tend to specialize in this overall offence type. Part IV summarises and discusses results across the three parts.

Part I starts with a paper dedicated to determining which offences ought to be counted as 'violent' for the purposes of risk prediction. Paper 1 has two steps. Assessment data on the content of violent index offences are analysed to identify the prevalence of violent behaviours, thus going beyond the information conveyed by the statutory charges for which offenders were convicted. Then, the associations between different types of reoffence and static and dynamic risk factors are explored. This leads to a classification which includes several statutory offence groups not usually considered violent in previous research, while excluding sexual offences. Paper 2 then formally constructs a logistic regression model to predict violent reoffending, using this offence classification. It validates the resultant OASys Violence Predictor on a separate sample, compares its predictive validity with that of other tools available for use in NOMS (including a presentation of sensitivity and specificity statistics, illustrating the practical effect of improvements in predictive validity), and tests different score weightings. Paper 3 compares the predictive validity of OVP and other risk tools specifically for nonsexual violent reoffending by sexual offenders, a topic of considerable interest to NOMS programme managers who wish to determine the content of

assessment batteries to identify offenders who are higher-risk for nonsexual as well as sexual recidivism.

Part II consists of Paper 4. This utilises a very large sample of OASys assessments and re-assessments, allowing changes in OVP dynamic risk factor scores to be tracked over time and related to subsequent violent reoffending outcomes. This provides strong evidence on whether the purportedly dynamic risk factors in OVP do change over time, whether they are causally related to reoffending, and whether accounting for changes in dynamic risk improves OVP's predictive validity and therefore increases the measured improvement associated with static/dynamic over static-only actuarial risk scales.

Part III consists of Paper 5. This examines the criminal careers of offenders with a known history of sexual reoffending. It looks at the extent of crossover between four classes of sexual offence, both in criminal histories and in reoffending. It also quantifies the extent of nonsexual violent and nonsexual nonviolent criminal histories and recidivism among offenders with various types of sexual offence history.

Part IV entails a general discussion of the thesis. It summarises the findings, and details the wider implications of this research.

Aims

The overall aim of this thesis is to promote better estimation of violent and sexual reoffending risk by constructing and validating an actuarial violence risk prediction scale, and testing for

the presence and impact of causal dynamic risk factors, and improving understanding of specialization in violent and sexual reoffending.

Specific aims

1. To empirically construct a classification of offences which are suitable for inclusion in a predictor of violent reoffending, which is based around a core of homicide and assault offences.
2. To construct and validate the OASys Violence Predictor (OVP), a new predictor of violent recidivism (with violence as classified above) for operational use by the National Offender Management Service.
3. To validate OVP for offenders with a history of sexual offending.
4. To examine sequences of offender assessments to measure changes in OVP scores and subscales, and determine whether these changes are associated with changes in the likelihood of violent recidivism.
5. To understand the extent to which sexual offenders specialize in particular types of sexual offence, and the extent to which they are specialists in sexual offending rather than generalist offenders.

Samples

All five chapters of the thesis used a common method for obtaining and processing data, with the addition of one other dataset in Chapter 1. However, they did not use identical datasets as the data sources were updated several times, and record selection was required for Chapters 3 and 5. Each chapter used the most contemporary data available at the time of its completion. As will become apparent, the chapters were not completed in chronological order.

Records were extracted from the Offender Assessment System research database maintained by the National Offender Management Service (NOMS). The OASys research database allowed a small number of follow-ups in limited areas of England and Wales in 2002, and “rolled out” to cover all English and Welsh probation services by early 2005. Data completeness and consistency filters were applied, and assessments completed near the start of a potential reoffending follow-up period (i.e., at the start of a community sentence or upon discharge from custody) were then selected. Personal identifiers for each offender remaining in the sample were submitted to the Ministry of Justice for matching with their Police National Computer (PNC) research/statistics database. Records for successfully matched offenders were processed to code criminal history variables, existing predictors of reoffending and statuses for recidivism variables involving various types of reoffending and durations of followup, including survival analytic followups in some instances. The resulting criminal record summaries were matched with the full OASys data to create datasets ready for analysis. Table 1, and the following text, summarise the assessment counts and date ranges used in each chapter’s analysis. Unless otherwise stated, 'assessments' listed below had been processed according to the above steps.

Table 1

Counts and date ranges of Offender Assessment System assessments, and dates of Police National Computer searches

Sample type and chapter	Number of initial assessments	OASys date range	PNC search date
All assessment (no PNC search)			
Chapter 1	230,334	Apr 2007 – Mar 2008	n/a
All offenders, with PNC search			
Chapter 1	26,619	Jan 2002 – Sep 2004	Nov 2008
Chapter 2 (construction)	15,918	Jan 2002 – Sep 2004	Nov 2008
Chapter 2 (validation)	49,346	Oct 2004 – Sep 2005	Nov 2008
Chapter 4	196,493	Oct 2004 – Mar 2009	Jul 2010
Sex offenders, with PNC search			
Chapter 3	21,445	Oct 2004 – Mar 2009	Jun 2012
Chapter 5	14,904	Oct 2004 – Mar 2008	Dec 2010

Chapter 1 utilised assessments completed between January 2002 and September 2004. It also utilised a set of OASys assessments completed between April 2007 and March 2008, which were filtered for data completeness and consistency but not matched with the PNC.

Chapter 2 also used many of the 2002-2004 assessments to construct the OASys Violence Predictor, and assessments completed between October 2004 and September 2005 to validate

it. (The Chapter 1 and 2 PNC search dates are the same because of concerns about the quality of an earlier PNC data search which had been used for preliminary Chapter 1 analysis.)

Chapter 4 used assessments completed between October 2004 and March 2008 to examine changes in OVP and its component items/scales over time. Chapters 3 and 5 selected offenders with an official criminal history of sexual offending or a current offence identified on OASys as having a sexual element or motivation, from assessments completed by March 2008 (Chapter 5) or March 2009 (Chapter 3, which was completed last).

The British Psychological Society code of ethical practice was adhered to in the design of research projects. Ethical approval was gained from the following: the Home Office, Ministry of Justice and National Offender Management Service (all for use of OASys data – its governance changed over the course of the five studies), the Police Information Approval Panel (for use of PNC data), and the University of Birmingham's School of Psychology Ethics Committee. Copies of these approvals are provided in Appendix A.

STATEMENT OF AUTHORSHIP

Chapters 1-5 contain material that has been published or awaiting publication in peer-reviewed journals². As such, each chapter has its own introduction and discussion. While repetition of material has been avoided where possible, the method sections include similar explanations which were replicated and adapted as necessary to ensure consistency across papers. The authorship on each article indicates collaborative working. To clarify, I am the senior author of all five papers. My supervisor Dr. Louise Dixon is a named author on three papers. My co-authors on the other two papers, Georgia Barnett, Helen Wakeling and Dr. Ruth Mann, are colleagues at the National Offender Management Service who provided advice on the analysis of data relating to sexual offenders.

² Chapter 1 - Howard, P. D., & Dixon, L. (2011). Developing an empirical classification of violent offenses for use in the prediction of recidivism in England and Wales. *Journal of Aggression, Conflict and Peace Research*, 3, 141-154.

Chapter 2 - Howard, P. D., & Dixon, L. (2012a). The construction and validation of the OASys Violence Predictor: Advancing violence risk assessment in the English and Welsh correctional services. *Criminal Justice and Behavior*, 39, 287-307.

Chapter 3 – Howard, P.D., Barnett, G.D., & Wakeling, H.C. (2013). Predicting nonsexual violent reoffending by sexual offenders: A comparison of four actuarial tools. *Legal and Criminological Psychology*. Advance online publication.

Chapter 4 - Howard, P. D., & Dixon, L. (2012b). Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys Violence Predictor. *Law and Human Behavior*, 37, 163-174.

Chapter 5 – Howard, P.D., Barnett, G.D., & Mann, R.E. (2013). Specialization in and within sexual offending in England and Wales. *Sexual Abuse: a Journal of Research and Treatment*. Advance online publication.

CHAPTER ONE

DEVELOPING AN EMPIRICAL CLASSIFICATION OF VIOLENT OFFENCES

This chapter develops the classification of violent offences which is used as the recidivism outcome for the OASys Violence Predictor. The use of empirical methods to form a class of reoffences to be predicted is novel but adds rigour to the predictor. The classification is founded upon examinations of, firstly, violent behaviours in index offences with different statutory classifications and, secondly, the associations between dynamic risk factors (OASys criminogenic need domains), static dynamic risk factors (including previous sanctions for each offence group) and recidivism for a range of offences. A set of offences to be considered “violent” for the purpose of recidivism prediction is therefore recommended.

To clarify the layout of Table IV, note that this reports the estimated beta coefficients of four stepwise logistic regression models, and the abbreviation NS indicates terms which were removed from these models as they were not statistically significant.

The following article was accepted for publication in the *Journal of Aggression, Conflict and Peace Research*, volume 3, pages 141-154, in 2011.

CHAPTER TWO

THE CONSTRUCTION AND VALIDATION OF THE OASYS VIOLENCE PREDICTOR (OVP)

This chapter constructs the OASys Violence Predictor (OVP), and validates it against a separate sample of offenders. A logistic regression model is fitted, and the results rounded, to create a robust and user-friendly predictor of violent reoffending. The validation stage includes testing of prediction of more serious violent outcomes, and detailed comparison with existing risk prediction tools, both of which confirm OVP's potential to improve risk assessment practice considerably. There are several reasons for the success of this exercise: the use of Chapter One's violence classification together with the wide range of dynamic risk factors covered by the Offender Assessment System and the very large sample sizes available from the national OASys research database.

The following article was accepted for publication in *Criminal Justice and Behavior*, volume 39, pages 287-307, in 2012.

CHAPTER THREE

PREDICTING NONSEXUAL VIOLENT REOFFENDING BY SEXUAL OFFENDERS: A COMPARISON OF FOUR ACTUARIAL TOOLS

This chapter principally tests whether OVP is an adequate predictor of nonsexual violent reoffending amongst offenders with a sexual offence history, and therefore whether NOMS can use it for this purpose. OVP is compared with Risk Matrix 2000/V and /C, which are predictors of nonsexual violent and combined sexual and violent reoffending designed for sexual offenders, and the general reoffending predictor OGRS3. Comparisons of the Area Under Curve predictive validity metrics for the predictors and their subscales are made, and further comparisons of three of the predictors are made after controlling for the different distributions of the predictors. The results confirm that OVP can be used with sexual offenders, as its predictive validity is equal to or exceeds that of other scales. Further analyses provide further insight upon risk prediction and specialisation among sexual offenders, by looking at the validity of OVP's component risk factors, and at rates of nonsexual violence among offenders with different types of sexual offence history.

The following article has been published in Online First form by Legal and Criminological Psychology.

PREDICTING NONSEXUAL VIOLENT REOFFENDING BY SEXUAL OFFENDERS: A COMPARISON OF FOUR ACTUARIAL TOOLS

Philip D. Howard ^{a,b}, Georgia D. Barnett ^a and Helen C. Wakeling ^{a,b}

^a National Offender Management Service, Ministry of Justice, UK.

^b School of Psychology, University of Birmingham, UK.

Author correspondence: Philip Howard, Planning and Analysis Group, National Offender Management Service, 5th Floor, Clive House, 70 Petty France, London SW1H 9EX.

philipdavidhoward@gmail.com

Abstract

Purpose. This study compared the ability of four risk assessment scales to predict nonsexual violent reoffending, and differences in nonsexual violent reoffending rates by sexual offending history.

Method. Risk assessment instruments were scored, and criminal histories and three nonsexual violent reoffending outcomes were coded, for a large sample of sexual offenders supervised by probation services in England and Wales. Predictive validities for the three outcomes were compared, varying the banding of risk scores to reflect practical constraints on offender management resources. Reoffending rates were compared by sexual offending history.

Results. After adjusting for risk assessment tool banding, the Offender Group Reconviction Scale version 3 and OASys Violence Predictor (OVP) had generally superior predictive validity to Risk Matrix 2000's V and C scales. However, several of OVP's dynamic risk factors were unrelated to nonsexual violent recidivism. Nonsexual violent reoffending rates were greater among those with prior but not current sexual offences and lower among those with indecent images offences, and sexual reoffending rates were lower but not negligible among those who had only sexually offended before the age of 16.

Conclusions. The use of OVP was recommended to the English and Welsh correctional services. The dynamic risk factor and sexual offence history results suggest that further work is required to optimise prediction of nonsexual violence among sexual offenders.

Introduction

Assessing the risk posed by offenders who have committed serious offences is a key task of any criminal justice system. Such assessment can form the basis for sentencing and parole decisions and for level and type of supervision, restriction and treatment. Through a process of empirical testing risk assessment has improved greatly over the last twenty years, evolving from unstructured, clinical judgement, to more structured schemes based on statistically or theoretically relevant factors that demonstrate a reliable relationship with recidivism. The risk assessment of sexual offenders continues to be particularly pertinent to the public, politicians, and those working in criminal justice. While establishing risk of sexual recidivism is a key aim of such assessment, studies following offenders in the community have suggested that it is also important to consider sexual offenders' risk of violent recidivism.

For example, Thornton and Travers (1991) found that, over a ten-year follow-up, a fifth of a national sample of sexual offenders released from prison in England and Wales were reconvicted of a nonsexual, violent offence; as many as were reconvicted of a sexual reoffence. Grubin (2008) reported a violent recidivism rate of 12.3% and a sexual recidivism rate of 10.8% over five-years for a national sample of convicted sexual offenders in Scotland. A recent meta-analysis examining the predictive accuracy of a static risk assessment tool across diverse, international sexual offender samples reported an average sexual recidivism rate of 14.9% over five years, compared to 24.3% violent (including sexual) recidivism rate in the same period (Hanson, Helmus & Thornton, 2010). As a result, it seems important that any assessment of sexual offenders' risk includes an assessment of their likelihood of being reconvicted for a violent offence, as well as their likelihood of being reconvicted for a sexual offence.

There are a range of risk assessment schemes that aim to do just this. The Risk Matrix 2000 (RM2000; Thornton et al., 2003) is an empirically-derived actuarial risk assessment tool, which is the

most commonly used tool to assess sexual offenders in England and Wales. It is currently the standard static risk assessment used by the Prison, Probation and Police services in these countries. It uses static information about offenders to classify them into risk bands that should differ substantially in their rates of reconviction for sexual and other nonsexual violent offences. The tool was developed for use in the United Kingdom with males aged 18 and over who have been convicted of a sexual offence committed after the age of 16. The RM2000 has three scales: The RM2000/s is a prediction scale for sexual recidivism, the RM2000/v a prediction scale for nonsexual, violent recidivism, and the RM2000/c is a combination of the first two scales and predicts sexual or nonsexual violent recidivism. The RM2000 was developed using a construction sample of 1,910 untreated convicted sexual offenders who had been discharged from prison in England and Wales, and who had been followed for two years (Thornton et al., 2003). The RM2000/s was constructed using existing research knowledge to identify individual factors predictive of recidivism to be incorporated into the tool and determine the weight to assign to each of these factors. Thornton and Travers' (1991) findings (as mentioned earlier) led to development of the RM2000/v and c scales.

The three RM2000 scales have been cross validated using further samples of sexual offenders (e.g., Thornton et al., 2003) with longer follow-up periods. However, these studies have been criticised (Grubin, 2008) for, among other things, having limited descriptions of the samples used, making it unclear how representative they were of the general sexual offender population. Two recent large-scale cross-validations of the tool have, however, provided more robust support for the use of this tool with English, Welsh and Scottish sexual offenders (Barnett, Wakeling & Howard, 2010; Grubin, 2008). The former study concluded that the RM2000 was a robust tool that had good predictive accuracy across a range of subgroups of offenders. However, the heterogeneity of sexual offenders continues to present difficulties in producing risk assessment instruments that are valid for all types of offender, and the designers of such instruments continue to advise that administrators use caution when using such assessments with marginal offender groups (e.g., Thornton, 2007).

While RM2000 is the most popular assessment for use with sexual offenders in the UK, practitioners in the National Offender Management Service (NOMS) of England and Wales are now able to use a range of actuarial measures when assessing a sexual offender's risk of nonsexual violent recidivism. As well as RM2000/v, they could use the Offender Group Reconviction Scale version 3 (OGRS3; Howard, Francis, Soothill & Humphreys, 2009) or the Offender Assessment System (OASys) Violence Predictor (OVP; Howard & Dixon, 2012). OGRS3 is a static actuarial predictor of proven reoffending for any offence, not just violent offences. OVP is a recently developed actuarial tool for predicting nonsexual violent proven reoffending, and combines static and dynamic risk factors. Neither OGRS3 nor OVP were designed specifically for sexual offenders, but are currently used within NOMS for other types of offender groups. The inclusion of dynamic risk factors for violent recidivism (as in OVP) is arguably an improvement over those tools that include only static factors, which are derived from data-driven approaches to recidivism prediction. Such approaches have helped identify those factors that are statistically related to recidivism, but not why these relationships exist. It is important that, alongside data-driven research, there is a focus on understanding the theoretical links between the predictors and violent offending.

To examine how well different tools fare in predicting nonsexual violent outcomes, Hanson and Morton-Bourgon's (2009) meta-analytic study examined the predictive accuracy of four of the most popular violence risk assessments used with sexual offenders, as well as examining how well assessments of risk of sexual recidivism fared in predicting violent outcomes. The best predictors of violent recidivism were those classed as empirical actuarial assessments that specifically assessed risk of violent or general recidivism, rather than sexual recidivism. Actuarial assessments are statistically derived tools which define ways of coding and combining historical information to produce a probability of the occurrence of a particular outcome within a specified time frame. Empirical actuarial tools are defined as those whose methods for item selection and/or combination are selected on the

basis that they evince an empirical relationship with recidivism (Hanson & Morton-Bourgon, 2009). Unstructured clinical judgement demonstrated the poorest predictive validity of violent recidivism while tools such as the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 2006), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 2006) and the Risk Matrix 2000-combined scale (Thornton et al., 2003) performed well (Hanson & Morton-Bourgon, 2009). A recent meta-analysis of studies of the predictive accuracy of Risk Matrix 2000 reported that RM2000/v predicted nonsexual violent recidivism with large effect size (Helmus, Babchishin & Hanson, 2013).

The OVP and OGRS tools were not included in Hanson and Morton-Bourgon's (2009) meta-analysis since they were developed and are primarily used only in England and Wales. (The other nations of the United Kingdom have separate criminal jurisdictions and risk assessment procedures.) Given their prominence in this jurisdiction, it is therefore pertinent that we examine these tools' effectiveness in predicting violent recidivism, compared with the 'incumbent' tool for this outcome among sexual offenders, RM2000/v. Given that the risk level of an offender is critical in determining their level of supervision, treatment pathway and ultimately their progression through the criminal justice system, it is important that the most robust risk prediction tools are used for each outcome of interest. Given the inherent resource constraints in the criminal justice system, and the desirability of not 'over-treating' low-risk sexual offenders (Wakeling, Carter & Mann, 2012), the risk prediction tools selected should ideally identify lower-risk offenders requiring a lesser degree of supervision and treatment, as well as the highest-risk offenders requiring maximal treatment and public protection resources. The ultimate aim of this research is to determine the best available tool in predicting nonsexual violent recidivism amongst sexual offenders in order to direct policy. The specific aims are to:

- 1) Compare the ability of four measures (OGRS3, OVP, RM2000/v, RM2000/c) in predicting nonsexual violent recidivism within a large sample of offenders with a known history of sexual offending.

Predicting nonsexual violent reoffending by sexual offenders

- 2) Understand whether nonsexual violent recidivism varies additionally according to offenders' histories of specific types of sexual offending.
- 3) Make recommendations based on the findings on the best tool to use to predict nonsexual violent recidivism amongst sexual offenders.

Method

Sample

The sample consisted of 21,445 known sexual offenders who commenced followup in the community (either as a result of release from prison or start of a community sentence) by the end of March 2009 and who had been assessed using OASys. The offenders in this study were extracted from a larger OASys dataset of 353,223 offenders, used previously for studies of general and violent reoffending. This larger sample already had proven reoffending (akin to reconviction) follow-up data as sought from the Ministry of Justice Police National Computer (MoJPNC) database. Filters had already been applied to the larger sample to ensure that all records had sufficiently complete and timely OASys and MoJPNC data quality to accurately score OVP, OGRS3 and RM2000/v. Of this OASys sample, 22,776 could be identified as sexual offenders, either because their MoJPNC records included, at any time, cautions (an alternative to prosecution issued for minor offences) and/or convictions for sexual offending, and/or because they had a *sexual element/motivation offence*. That is, OASys was used to identify that their current conviction had a sexual element or motivation (OASys questions 2.2F and 2.9 respectively) despite the statutory offence being nonsexual in nature. A group of 1,331 offenders were removed because their reoffending followup commenced by September 2004, and they had therefore been included in the original OVP construction sample (Howard & Dixon, 2012b) and should therefore not be used to validate it. This left 21,445 offenders who had sufficient information to score the tools and could be identified as sexual offenders. Of those 21,445, 13,040 been 'at risk' for at least four years when the proven reoffending data was sourced (i.e., at least

four years had passed between the start of their community sentence or release from prison and the cutoff date for proven reoffending, six months before the MoJPNC update on 29 June 2012). Of these 13,040, note that 1,139 (9%) are among the group for whom RM2000 was not originally constructed, because their sexual offending history was confined to offences committed before the age of 16. We include these offenders in order to investigate the predictive validity of different tools for this group, for whom no one standard risk assessment is used. The four-year follow-up period was chosen through scrutiny of the numbers of offenders eligible for different lengths of follow-up and the numbers committing each of the three types of nonsexual violent reoffence (see Tables 2 onwards). Statistical power is maximised when Npq is maximised, N being the sample size (at this length of followup) p being the proportion reoffending and q being $(1-p)$ (Harrell, Lee & Mark, 1996). Despite some variation across the three reoffending outcomes, there was a clear overall tendency for Npq to be maximised with a four-year follow-up.

The sample, while not including all sexual offenders serving or finishing sentences in the community or discharged from custody during this time, is likely to be fairly representative of this population. Omission from the OASys database would require the offender to have both received only a limited report prior to sentence and to have been sentenced to a custodial sentence of less than 12 months or a community sentence involving neither supervision nor treatment. This is likely to be relatively rare for those with histories of sexual offending. Therefore, the sample is not preselected on the basis of risk as nearly all sexual offenders should be assessed using OASys. This sample thus represents what Hanson, Helmus, and Thornton (2010) term a ‘routine’ correctional sample, rather than a preselected high-risk sample.

Table 1 lists the demographic characteristics of the sample. The vast majority of the sample was White, and most (62%) had an index offence relating to sexual crime. Those who had a nonsexual offence as their index offence had a history of sexual offending (defined as at least one previous conviction or caution for a sexual offence). The most common sexual offence committed by those in

Predicting nonsexual violent reoffending by sexual offenders

the sample was sexual assault, followed by possession or manufacture of indecent images of children (usually relating to internet-based offences). However, about one-fifth of current sexual offences (12% of the total sample) were sexual element/motivation offences. The majority of these were statutory nonsexual violent offences. Offenders who had been followed up for less than four years were less likely to have a nonsexual violent index offence but more likely to have an acquisitive index offence.

Predicting nonsexual violent reoffending by sexual offenders

Table 1. Sample demographic characteristics

		4-year follow-up sample		Cases followed up for less than 4 years	
Variable		n	M (SD)	n	M (SD)
Age		13,896	37.7 (13.5)	9,002	38.8 (13.8)
		n	%	n	%
Ethnicity					
	White	10,888	78.4	6,714	74.6
	Black	705	5.1	498	5.5
	Asian	481	3.5	308	3.4
	Mixed	199	1.4	130	1.4
	Other	82	0.6	59	0.7
	Missing	1,541	11.1	1,293	14.4
Index Offence Type					
	Rape/buggery (child victim)	527	3.8	342	3.8
	Rape/buggery (adult victim)	560	4.0	367	4.1
	Sexual assault (child victim)	1,360	9.8	727	8.1
	Sexual assault (adult victim)	1,176	8.5	729	8.1
	Unlawful sexual intercourse with child under 13 or 16	687	5.0	574	6.4
	Gross indecency with children	166	1.2	68	0.8
	Other contact sexual offences	246	1.8	218	2.4
	Indecent images of children	1,156	8.3	898	10.0
	Indecent exposure	434	3.1	322	3.6
	Other noncontact sexual offences	11	0.1	18	0.2
	Noncompliance with requirements of sex offender orders	330	3.2	375	4.2
	Nonsexual statutory offence with sexual element or motivation	1,596	11.5	901	10.0
	Violent offences (according to RM2000/V definition)	1,985	14.3	946	10.8
	Violent offences (according to OVP but not RM2000/V definition)	1,049	7.6	459	5.1
	Acquisitive offences	1,404	10.0	1,301	14.5
	Other	1,469	10.6	937	10.4

Note. The classification of violent offences used for RM2000/v includes offences of homicide and nonsexual assault, robbery, aggravated burglary, cruelty/neglect of children and animals, and possession of firearms. OVP

Predicting nonsexual violent reoffending by sexual offenders

also classifies other weapon possession offences, threats and harassment, criminal damage and public order offences as violent. Sexual element/motivation offences identified from OASys questions 2.2F ('Did any of the [current] offence(s) involve a sexual element?') and 2.9 ('Evidence of sexual motivation'); sexual element/motivation was counted where the index offence type would otherwise have been from one of the rows below this on the table.

Measures

Risk Matrix 2000 (RM2000)

RM2000 (Thornton, 2007) is a static risk assessment tool for use with adult males who have been convicted of a sexual offence. At least one of the sexual offences must have been committed when the offender was over 16. Due to the limitations of OASys data, this paper uses a modified version of the sexual recidivism subscale, RM2000/s. Barnett et al. (2010) describe in detail how this version differs from full RM2000/s, and estimate that it has slightly reduced predictive validity for sexual recidivism, with an AUC of 0.69 rather than 0.71. This modified version is made up of five items divided into two scoring steps. Step one comprises three items: age of the offender on release from custody or current age in the community, number of sentencing occasions for a sexual offence and number of sentencing occasions for any criminal offence. The sum of scores for these items is translated into one of four preliminary risk categories: Low, medium, high or very high. In the second step, the risk category is raised one level if the offender has ever had male sexual offending victims and/or been convicted of noncontact sexual offences. The second step is not applied to offenders whose sexual offending history is entirely internet-based.

RM2000/v, which predicts violent recidivism, is composed of three items; age, number of sentencing occasions for a violent offence, and whether or not the offender has ever been convicted of a burglary. The sum of scores for these items is translated into one of the four risk categories described above. The RM2000/c combines the scores from the s and v scales to produce an overall risk classification that predicts sexual *and* nonsexual violent recidivism.

RM2000/v classifies violent offences as those “whose legal definition implies the use or threat of force against the person but [not] that solely imply sexual violence” (Thornton, 2007, p.26). These ***RM2000/v-class reoffences*** include homicide/assault, threats and harassment offences, robbery/aggravated burglary, arson and abduction. In operational use, cruelty to animals and

possession of a firearm are to be considered on a case-by-case basis (considering, respectively, whether the cruelty was non-sexual and whether the context of firearm carrying implied possible use against a person). For the purpose of this study, all animal cruelty offences were counted as violent, and firearm carrying offences were counted as violent if the precise legal charge implies actual or threatened use against a person.

Offender Group Reconviction Scale 3 (OGRS3)

OGRS3 (Howard et al., 2009) is an actuarial predictor of general reoffending, used extensively in assessments completed by probation and prison staff in England and Wales. It estimates the percentage likelihoods of proven reoffending (any conviction or caution for a new offence) committed within 1 and 2 years of the start of a community sentence or discharge from custody. It includes only static risk factors: age, sex and criminal history. The criminal history factors include a 20-group classification of current offence type, an offending ‘rate’ based on the number of sanctions (convictions and cautions) in the offender’s criminal history and the number of years between first and current sanction, and an additional variable identifying those with very short criminal histories. Despite its exclusively static content, and design as a predictor of general reoffending, OGRS3 achieves moderate levels of predictive validity for nonsexual violent reoffending, with an AUC of .70 when tested with a sample of almost 50,000 NOMS offenders (Howard & Dixon, 2012a).

OASys Violence Predictor (OVP)

The OASys Violence Predictor (OVP) predicts proven reoffending over a broad range of nonsexual violence-related (“*OVP-class*”) offences: homicide/assault, threats/harassment, criminal damage, public order, robbery/aggravated burglary and weapon possession. This range of offences was decided upon through an empirical process which considered how each offence subclass was related to assessor notes of violent index offence content (e.g., weapon use, excessive/sadistic violence), and

Predicting nonsexual violent reoffending by sexual offenders

associations with static and dynamic risk factors and nonsexual violent reoffences of varying severity (Howard & Dixon, 2011). OVP uses a 100-point scoring system to combine information on static risk factors (age, sex and previous sanctions for violent and nonviolent offences, totalling 60 points) and dynamic risk factors (accommodation, employment, alcohol misuse, psychiatric treatment, temper control and two attitudinal measures, totalling 40 points). The 100-point scores are translated into one- and two-year predicted percentages for violent reoffending. Howard and Dixon (2012a) describes its construction and validation on separate samples of general OASys offenders. It was validated as a predictor of not only OVP-class offences but also as a predictor of homicide and assault and the most serious nonsexual violent offences (i.e., homicide and wounding with intent to cause grievous bodily harm). On this general sample, it was found to be a better predictor of all violent outcomes than the now obsolete “OASys score” (see below), OGRS3 and RM2000/v. It also displayed high predictive validity in a study of later assessments, which provided evidence that it includes causal dynamic risk factors (Howard & Dixon, 2012b).

The Police National Computer (PNC) research database

The Police National Computer (PNC) is the operational system used by all 42 police forces in England and Wales to record details of suspected and proven offenders, as well as details of crimes solved and under investigation. The Ministry of Justice’s PNC research database contains extracts of criminal records data on cautioned and convicted offenders. It is available to researchers through the Ministry of Justice’s Analysis and Statistics group. It is the source of data on previous sanctions and proven reoffending.

Offender Assessment System (OASys)

The Offender Assessment System (OASys; Home Office, 2006) is a structured clinical risk/needs assessment and management tool, used throughout NOMS to inform court reports and the

Predicting nonsexual violent reoffending by sexual offenders

management of offenders serving community and custodial sentences. It consists of four main components: An analysis of offending-related factors, a risk of serious harm analysis, a summary sheet and a sentence plan. The offending-related factors component includes 13 sections which cover criminal history, analysis of [current] offences, assessment of ten dynamic risk factors and suitability to undertake sentence-related activities (e.g., unpaid work, offending behaviour programs). Each of the dynamic risk factors is assessed using between four and ten questions, each scored on a 0/2 or 0/1/2 basis. The ranges of total scores on the dynamic risk factors therefore run from 0-8 to 0-20. These factors can be used in the prediction of reoffending, and assist the assessor in developing and reviewing the offender's Sentence Plan. The offending-related factors component is the source of dynamic risk factor data used to score OVP. The risk of serious harm analysis component provides a structure for clinical case formulation and Risk Management Plan for offenders considered likely to commit harmful acts in the future. The summary sheet component combines scores from each of the offending-related factors to produce predictive scores. Howard and Dixon (2012) describe the construction and validation of OVP, which was introduced into OASys practice in August 2009. While OASys did not specifically assess offenders' risk of violent recidivism until August 2009, the questions used to construct OVP were included in OASys prior to this date, and OVP scores can therefore be computed in the present study. (RM2000/v was used to assess sexual offenders before this date, but it was scored separately from OASys.)

OASys is used throughout NOMS with offenders aged 18 and over who are convicted awaiting sentence, serving custodial sentences of at least 12 months, or serving probation sentences involving supervision. Assessments are reviewed periodically over the course of the sentence. In 2010/11, approximately 860,000 assessments were completed on 360,000 offenders by 18,500 staff. Data from completed assessments are copied to a central research and statistics office within NOMS headquarters. Data completeness and integrity checks are undertaken before producing subsets for analysis, such as the current samples.

Procedure

OGRS3 and RM2000/v scores were calculated from Police National Computer (PNC) data. OVP scores were calculated from a combination of PNC and OASys data. RM2000/s scores, as a component of RM2000/c, were simulated using the procedure described by Barnett et al. (2010).

Each tool's score is presented in multiple forms. RM2000/v and c are usually presented in categorical form (Low / Medium / High / Very High; L/M/H/VH). OGRS3 is usually reported as a two-year percentage score, and as OVP as a two-year percentage and a 0-100 point score. However, we also present: RM2000/v as an uncategorised 0-8 point scale, and RM2000/c as an uncategorised 0-6 point scale; OVP's static (0-60) and dynamic (0-40) subscales, and OGRS and OVP L/M/H/VH categories. The OGRS and OVP categories are reported to NOMS staff on the OASys summary sheet, and have previously only been published internally (NOMS, 2009).

Proven reoffending was the outcome of interest for this study. Proven reoffending comprises offenses committed within 48 months of the date of community sentence or release from custody, which have led to a formal criminal sanction no more than six months after the end of the follow-up period. AUCs were produced for three 48-month proven reoffending outcomes: any offence included in the RM2000/v classification; any offence included in the OVP classification; and ***homicide and wounding*** (i.e., murder, attempted murder, manslaughter, threats to kill, wounding with intent to do grievous bodily harm and other violent acts of equivalent severity). The AUC indicates the accuracy of a scale's discrimination between recidivists and nonrecidivists; it can range from 0 to 1, where 0.5 represents prediction no better than chance, and 1 indicates perfect prediction. Rice and Harris (2005) use comparisons to Cohen's *d* to suggest that an AUC value of 0.71 represents a large effect size and an AUC value of 0.64 represents a moderate effect size. A robust nonparametric test for correlated measures (DeLong, DeLong & Clarke-Pearson, 1988) was used to compare AUCs. For reference, we also present sexual reoffending rates, though analysis of AUCs is not undertaken for this outcome.

The use of AUCs was especially important as the distributions of banded scores varied considerably between the four tools. Restriction of range through banding reduces relative risk, which means that placing people into fewer risk groups on a tool will result in poorer AUCs (see also Hanson, 2008). However, in practice, using risk bands is necessary so that actuarial tools clearly identify offenders at high or low risk, allowing operational staff to direct scant resources towards those representing the greatest risk of serious reoffending. In order to test how great a difference restriction of range (as a result of banding offenders into fewer risk categories) makes, both the total and banded scores of each risk tool were examined. (The exceptions to this were the OVP static and dynamic subscales, as these do not have a banding system.) To completely control for the effect of risk distribution, a further set of predictive validity comparisons were conducted, using groups of ordered OGRS3 and OVP 0-100 total scores, which match the distribution of the RM2000/v 0-8 scores.

RM2000/v was designed for use only with offenders who had committed at least one sexual offence at age 16 or over, whereas NOMS must manage offenders with convictions of sexual offending at any age, some of whom must be managed under public protection arrangements (National Offender Management Service, 2012). Also, some previous research studies only consider offenders with a current sexual offence conviction, though all four tools are designed to be used with those with prior sexual offence convictions. Therefore, comparisons of the tools' predictive validity were made for those with and those without sexual offences committed when they were 16 or over, and for offenders with current and those with only prior sexual offence convictions. Offenders' history of indecent image offending was also considered in comparisons, as previous research on a related sample of offenders shows very different nonsexual reoffending patterns

As a consequence of the predictive validity results for the OVP static and dynamic subscales, a set of logistic regression models were created to identify the contribution of each of OVP's risk factors to prediction of the three nonsexual violent reoffending outcomes.

Finally, distributions of OVP risk bands and OVP-class nonsexual violent recidivism rates are compared for offenders with different sexual offence histories. *Four sexual offence type groups* were first defined. *Contact child* offences involve physical contact offending where the victim is known to be a child, either from the statutory offence code or from OASys victim information. *Contact adult* offences involve physical contact offending where the victim is known to be an adult or, in a small number of cases, is of unknown age. (We avoid the term ‘rapist’ to describe those in our sample who have committed contact adult offences, as English and Welsh statutory offences include ‘rape of a girl aged under 16’ and similar offences.) *Indecent images* offences involve the making, distribution, showing and advertisement of indecent images of children. Indecent images usually involve the internet, but this is not inherent in the offence, and indeed statutorily-defined grooming offences, which typically involve the internet, are included in the contact child group, as the motivation of this offence is to make sexual contact with a child. The remaining group, *paraphilia*, includes offences resulting from a range of sexual interests that are usually most easily gratified through criminal behaviour; victims may be of any age. Most paraphilia offences which result in criminal sanction are prosecuted as indecent exposure (i.e., exhibitionism) and are therefore noncontact offences, while some of the other offences in the group are related to voyeurism and zoophilia.

Results

Table 2 shows the four-year proven reoffending rates by violent and sexual offence outcome and Offending Group Reconviction Scale (OGRS) 3, OASys Violence Predictor (OVP), RM2000/combined sexual and nonsexual violence scale (RM2000/c) and violence scale (RM2000/v) risk categories. For all offenders in the sample, the rate of proven violent offending is slightly higher, and the rate of proven sexual offending slightly lower, than that reported in other studies that have used samples from within the U.K. (e.g., Grubin, 2008; Thornton et al., 2003). This is likely to be the

result of the inclusion in the current sample, of those whose index offence is nonsexual; such individuals were not present in Thornton et al.'s (2003) nor Grubin's (2008) Risk Matrix 2000 validations. As those studies excluded offenders with nonsexual index offences, their samples are likely to have consisted of a higher proportion of individuals who specialize in sexual offending: that is, people whose every offence is sexually motivated. Conversely, the current sample is likely to consist of a higher proportion of generally criminal offenders, for whom sexual offending is part of a criminal lifestyle.

The distribution of offenders across the four risk bands (from low to very high) varies markedly across the different tools. Both OVP and OGRS classed the majority of the sample as low risk, while RM2000/c and v classed fewer offenders as low risk, and more as high and very high risk. Only 173 offenders were placed in the very high risk category according to the OVP, compared with over ten times that amount using the RM2000/v ($n = 2,157$). Rates of proven reoffending, regardless of violent offence outcome, increased across each ascending risk category for all tools used. That is, lower risk offenders had lower rates of proven reoffending than higher risk offenders. Rates were closest together for the rare outcome of homicide and wounding, for OGRS3 and OVP's high and very high categories and RM2000/v and RM2000/c's low and medium categories. Chi-square tests indicated that for all tools, for RMV-class offences, rates of offending between risk groups differed significantly ($p < .001$), as did rates of OVP-class reoffending ($p < .01$). For homicide and wounding offences, rates of reoffending differed significantly ($p < .01$) between all tools' risk groups, with the exception of OVP's high and very high risk groups ($\chi^2 = 0.74, p = .39$).

Table 2. Four-year proven reoffending rates of offenders in the different OGRS3, OVP and RM2000/v risk categories by nonsexual violent offence and sexual offence outcome

		Homicide and wounding	RM2000/v class offence	OVP class offence	Sexual offence
	<i>N</i>	% (<i>n</i>) reoffending	% (<i>n</i>) reoffending	% (<i>n</i>) reoffending	% (<i>n</i>) reoffending
All offenders	13,896	1.5% (209)	25.7% (3,569)	39.1% (5,426)	3.6% (505)
OGRS3					
Low	8,553	0.5% (41)	11.1% (948)	19.1% (1,639)	3.2% (271)
Medium	3,425	2.5% (87)	42.7% (1,461)	64.4% (2,206)	4.6% (156)
High	1,702	4.2% (71)	58.8% (1,001)	81.2% (1,382)	3.9% (67)
Very High	216	3.7% (8)	73.6% (159)	92.1% (199)	5.1% (216)
OVP					
Low	7,801	0.4% (33)	9.5% (738)	16.2% (1,266)	2.9% (229)
Medium	4,452	2.4% (107)	41.0% (1,823)	62.6% (2,787)	4.1% (183)
High	1,363	3.9% (53)	59.1% (805)	82.3% (1,122)	5.1% (69)
Very High	280	5.0% (14)	72.5% (203)	89.6% (251)	8.7% (24)
RM2000/v					
Low	3,702	0.1% (4)	3.6% (132)	7.2% (268)	2.6% (96)
Medium	3,995	0.6% (22)	15.5% (618)	27.4% (1,096)	4.4% (175)
High	3,663	2.1% (77)	38.2% (1,398)	58.3% (2,136)	3.7% (135)
Very High	2,536	4.1% (104)	56.0% (1,421)	76.0% (1,926)	3.9% (99)
RM2000/c					
Low	1,949	0.0% (0)	2.1% (41)	4.3% (83)	1.9% (36)
Medium	4,685	0.5% (22)	12.9% (602)	22.5% (1,055)	2.9% (137)
High	5,888	2.1% (123)	36.1% (2,124)	55.0% (3,239)	4.3% (251)
Very High	1,374	4.5% (62)	58.4% (802)	76.4% (1,049)	5.9% (81)

Note. OGRS3 risk categories are based on 2-year predicted probabilities of proven reoffending for any offence, as follows. Low: 0 – 49%; Medium: 50 – 74%; High: 75 – 89%; Very High: 90 – 99%. OVP risk categories are based on 2-year predicted probabilities of proven reoffending for OVP-class offences, as follows. Low: 0 – 29%; Medium: 30 – 59%; High: 60 – 79%; Very High: 80 – 99%. (NOMS, 2009.)

Table 3 presents the numbers of offenders with each combination of current/noncurrent sexual offender status, history of indecent image offending and age at last sexual offence, and reoffending rates for each of the four offence classes. These combinations were chosen on the basis of the importance of understanding whether predictors can be applied to noncurrent sexual offenders or those who had only sexually offended when aged under 16, and the evidence on lower nonsexual reoffending rates among indecent image offenders (Howard, Barnett, & Mann, in press). Four of the eight subgroups formed by these combinations were populated by very few offenders, or none at all, as it was rare for noncurrent offenders to have a history of indecent image offending or for current offenders to have committed their index offence when aged under 16. For all three violent outcomes across the remaining four subgroups, reoffending rates were higher among noncurrent offenders, especially those who had not offended sexually as adults (*noncurrent juvenile* offenders) rather than those who had offended sexually as adults (*noncurrent adult*). They were lower among current offenders without indecent image offences (*current nonimage*) and lowest among indecent image offenders (*current image*). The sexual reoffending rate was considerably lower among the noncurrent juvenile group but, allowing for variations in length of follow-up, still considerably higher than the base rates for a range of populations with no known history of sexual offending (Harris & Hanson, 2012).

Table 3. Comparison of four-year proven nonsexual violent reoffending and sexual offence outcomes by current sexual offending, indecent image offending and age at last sexual offence.

Sexual offender status	Any history of indecent image offending?	Age at last sexual offence	<i>N</i>	Homicide and wounding	RM2000/v class offence	OVP class offence	Sexual offence
				% (<i>n</i>)	% (<i>n</i>)	% (<i>n</i>)	% (<i>n</i>)
				reoffending	reoffending	reoffending	reoffending
All offenders			13,896	1.5% (209)	25.7% (3,569)	39.1% (5,426)	3.6% (505)
Current	No	Under 16	29	3.4% (1)	31.0% (9)	55.2% (16)	3.4% (1)
		16 or older	6,519	0.8% (51)	15.0% (981)	24.5% (1,597)	3.9% (254)
	Yes	Under 16	NIL				
		16 or older	1,154	0.3% (3)	2.4% (28)	4.3% (50)	3.7% (43)
Prior but noncurrent	No	Under 16	1,109	3.5% (39)	51.2% (568)	70.9% (786)	1.4% (16)
		16 or older	4,990	2.2% (111)	39.4% (1,968)	59.0% (2,945)	3.7% (183)
	Yes	Under 16	1	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
		16 or older	94	2.1% (2)	16.0% (15)	34.0% (32)	8.5% (8)

Figure 1 illustrates the varying distributions of RM2000/v score between offender subgroups. The variation was considerable, with 98% of current image, 73% of current nonimage, 32% of noncurrent adult and 14% of noncurrent juvenile offenders being rated Low or Medium risk.

Figure 1

Risk Matrix 2000/V score distribution of selected offender groups, by current/prior sexual offending, age at most recent sexual offence and history of indecent image offending

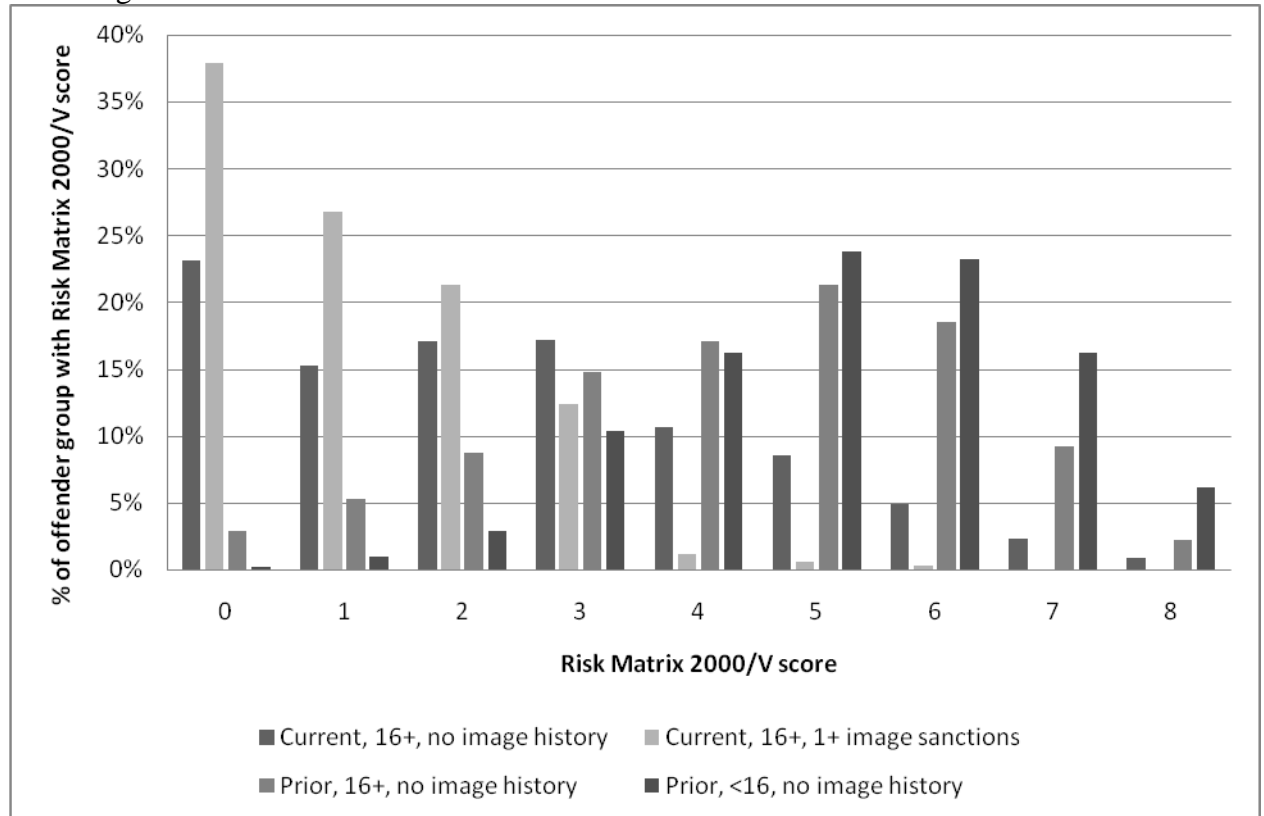


Table 4 illustrates how reoffending rates varied by risk across the four more frequent sexual offender subgroups, using RM2000/v categories only. For all four subgroups, nonsexual violent reoffending rates were associated with RM2000/v risk category. However, as in Table 3, rates were lower for current nonimage offenders and lower still for current image offenders, even after controlling for risk category. For example, RM2000/v-class reoffending rates among Medium risk offenders were 3% for current image offenders, 11% for current nonimage offenders, 26% for noncurrent adult offenders and 28% for noncurrent juvenile offenders.

Table 4. Four-year proven nonsexual violent reoffending rates for selected sexual offender subgroups, by Risk Matrix 2000/v risk band and nonsexual violent offence and sexual offence outcome

Offender type		Homicide and wounding	RM2000/v class offence	OVP class offence	Sexual offence
	<i>N</i>	% (<i>n</i>) reoffending	% (<i>n</i>) reoffending	% (<i>n</i>) reoffending	% (<i>n</i>) reoffending
All offenders	13,772	1.5% (209)	25.7% (3,569)	39.1% (5,426)	3.6% (505)
Current sexual offence aged 16+, no history of indecent images offences					
Low	2,505	0.1% (2)	2.6% (66)	5.8% (146)	2.3% (57)
Medium	2,238	0.4% (8)	11.3% (254)	22.0% (492)	4.8% (108)
High	1,249	1.5% (19)	31.3% (391)	49.1% (613)	4.6% (58)
Very High	527	4.2% (22)	51.2% (270)	65.7% (346)	5.9% (31)
Current sexual offence aged 16+, 1+ indecent images sanction					
Low	744	0.0% (0)	1.1% (8)	2.3% (17)	3.2% (24)
Medium	387	0.3% (1)	3.4% (13)	5.9% (23)	4.7% (18)
High	20	5.0% (1)	30.0% (6)	40.0% (8)	5.0% (1)
Very High	3	33.3% (1)	33.3% (1)	66.7% (2)	0.0% (0)
Noncurrent sexual offence, most recent sexual offence when aged 16+, no indecent image history					
Low	409	0.2% (1)	12.5% (51)	23.5% (96)	2.9% (12)
Medium	1,176	0.9% (10)	25.9% (305)	42.6% (501)	4.0% (47)
High	1,916	2.2% (42)	40.8% (781)	62.8% (1,203)	3.5% (68)
Very High	1,489	3.9% (58)	55.8% (831)	76.9% (1,145)	3.8% (56)
Noncurrent sexual offence, most recent aged <16, no indecent image history					
Low	13	7.7% (1)	30.8% (4)	30.8% (4)	NIL
Medium	147	2.0% (3)	27.9% (41)	46.3% (68)	NIL
High	444	2.9% (13)	47.1% (209)	66.0% (293)	1.1% (5)
Very High	505	4.4% (22)	62.2% (314)	83.4% (421)	2.2% (11)

Note. Offenders in the four least frequent subgroups in Table 3 (total *n* = 124) are not included in this table.

Given the varying distributions of the four predictors, it is difficult to judge their relative merits. RM2000/c has the lowest recidivism rates among its Low risk offenders, but places fewest offenders in this category. Conversely, the rates for OGRS3 and OVP's High and Very High risk groups are greater than those of RM2000/v and RM2000/c's Very High risk groups, but the latter groups contain more offenders.

Table 5 therefore shows the AUCs of each of the tools at the four-year follow-up. All tools demonstrated good predictive accuracy, with AUCs varying by around five points per outcome type when moving from unbanded to banded scores. AUCs were generally slightly higher when the outcome of choice was OVP-class reoffending, suggesting that less serious violent reoffences are easier to predict.

Table 5. Comparison of Area Under the Curve (AUC) statistics for four-year proven nonsexual violent reoffending outcomes for different violence risk assessments (using both unstandardised risk bands and total risk scores)

Measure	Range of possible scores	AUC (95% CI) by four-year proven reoffending outcome		
		Homicide & wounding	RM2000/v-class offences	OVP-class offences
OGRS3: banded	0 – 3	.73 (.70, .76) ♦♦♦	.75 (.74, .76) ♦♦♦	.77 (.76, .78) ♦♦♦
OGRS3: total	0 – 100	.77 (.74, .79)	.81 (.80, .82) ***	.85 (.84, .86) ***
OVP: banded	0 – 3	.73 (.70, .76) ♦♦♦	.76 (.75, .77) ♦♦♦	.79 (.78, .81) ♦♦♦
OVP: total	5 – 100	.77 (.75, .79)	.81 (.80, .82) ***	.85 (.84, .86) ***
OVP: static scale	5 – 60	.78 (.75, .80)	.82 (.82, .83) ***	.86 (.85, .86) ***
OVP: dynamic scale	0 – 40	.68 (.64, .71) ♦♦♦	.70 (.69, .71) ♦♦♦	.73 (.72, .74) ♦♦♦
RM2000/v: banded	0 – 3	.76 (.74, .79) ♦♦♦	.78 (.77, .79) ♦♦♦	.80 (.79, .81) ♦♦♦
RM2000/v: total	0 – 8	.78 (.75, .80)	.79 (.78, .80)	.81 (.81, .82)
RM2000/c: banded	0 – 3	.73 (.71, .76) ♦♦♦	.74 (.73, .75) ♦♦♦	.76 (.75, .76) ♦♦♦
RM2000/c: total	0 – 6	.74 (.72, .77) ♦♦	.76 (.75, .76) ♦♦♦	.78 (.77, .79) ♦♦♦

Note. $N = 13,896$. The OVP scoring scheme from Standard OASys was used. OVP awards five points for being male, so the minima of OVP's total and static scale among this sample are five rather than zero. ♦ denotes scores that are significantly worse than RM2000/v (total) using T-test comparisons, while * denotes scores that are significantly better than this tool.

♦ or * $p < .05$

♦♦ or ** $p < .01$

♦♦♦ or *** $p < .001$

All tools' banded and unbanded AUCs were compared to the RM2000/v (total) AUC, as this is the tool routinely used in the UK. (At this stage, no adjustment had been made for the different distributions of the banded versions of each tool.) All tools' total scores fared equally well in predicting homicide and wounding, with the exception of the RM2000/c which was significantly worse than the RM2000/v. The banded versions of all tools were inferior to the total scores, in accordance with the fact that restriction of range impacts negatively on AUC. For both RM2000/v-class and OVP-class violent reoffences, OGRS3 and OVP total scores were superior to the RM2000/v total score, while the RM2000/c total score was inferior; again, all banded scores were inferior to the RM2000/v total. The best predictor of these outcomes was the OVP static subscale, illustrating that the inclusion of the moderately predictive OVP dynamic subscale reduced OVP's overall predictive validity. (This topic is explored further below.)

Additional comparisons were made to check the predictive validity of each tool among the current and noncurrent sexual offence groups, and in the noncurrent adult and noncurrent juvenile subgroups. For RM2000/v-class reoffending, using RM2000/v category as the predictor, the AUC was 0.81 for the current group, 0.66 for the noncurrent group as a whole, 0.62 for the noncurrent juvenile subgroup and 0.66 for the noncurrent adult subgroup. Other comparisons are available on request from the authors. The AUCs for all tools varied between these groups and subgroups due to distributional issues: the current group had strong AUCs as it was able to contrast some higher-risk offenders with a large number of offenders with very low risk (scores of 0 on the 0-8 RM2000/v scale, as shown in Figure 1, and similarly low OGRS and OVP scores), while the noncurrent juvenile subgroup's weak AUCs were related to raised, compressed risk distributions on all tools.

The poorer performance of the banded versions of the OGRS3 and OVP could be a consequence of the fact that the bandings for these tools were not devised using sexual offender populations, with which the RM2000 scales were developed. Both OGRS3 and OVP bandings were developed using a general offender population, and their very high risk categories were intentionally designed to include very few offenders, to restrict to a manageable size, the number of offenders who would be directed to the highest tier of offender management in the community. Of the current sample, only 274 received a Very High Offender Assessment System (OASys) risk of serious harm rating (a clinical judgement following consideration of all available information, including actuarial predictions of violent and sexual reoffending), reflecting the very targeted use of this designation due to the limited resources available in practice. Only 4,040 were rated as High risk of serious harm. Thousands of offenders were rated Very High risk of reconviction on RM2000/v or RM2000/c. Should this rating translate to an OASys classification of Very High risk of serious harm, there would be serious resource implications.

While OGRS3 and OVP may therefore use risk bands with more practical appeal, the limited size of their top band effectively results in further restriction of range. Similar, among total scores, OGRS3 and OVP have longer ranges than RM2000/v or RM2000/c, which could create a bias towards them in total score comparisons. In order to determine whether differences in the predictive accuracy of the different tools were artefacts of the different distribution of offenders across risk bands, the next stage of analysis standardised the risk bands of OGRS3 and OVP to match the distribution of the RM2000/v 0-8 point total score. RM2000/c was not considered in this or further comparisons due to its clearly weaker performance in Table 5.

Table 6 depicts the marginal reoffending rates and AUCs of the tools when standardised to the RM2000/v banding. When the risk groups in each tool were evenly distributed, there was no significant difference between the three tools for homicide and wounding reoffending, while OGRS3 and OVP significantly outpredicted RM2000/v for the other two reoffending types. They had very similar predictive validity to one another; for OVP-class reoffending, the rounding AUCs differ by one point but the exact AUCs were 0.846 for OGRS3 and 0.844 for OVP. These results also provide a practical illustration of what certain AUC differences actually look like, addressing a common complaint the authors have heard from practitioners and clinician-researchers that AUCs are too abstract to be useful. The results also illustrate the differences in predictive validity which can exist when all AUCs are within Rice and Harris (2005)'s high predictive validity band. The one-point difference in RM2000/v-class prediction is visible as a series of small differences: RM2000/v had somewhat higher reoffending rates in the score-0 and score-1 bands and was inconsistently weaker at the top end (with lower rates at scores 6 and 8 outweighing higher rates at score 7). The three- and four-point differences in OVP-class violent offending prediction had a clearer effect: score-0, -1 and -2 reoffending rates for RM2000/v were 2 to 4 points higher than for OGRS3 or OVP, and rates at scores 5 to 8 were 2 to 10 points lower for RM2000/v than the other two predictors.

OGRS3 and OVP were also grouped into L/M/H/VH bands which matched the distribution of RM2000/v's bands. For homicide and wounding reoffending, there were no significant differences (AUCs for RM2000/v, OGRS3 and OVP were 0.77, 0.76 and 0.76 respectively; comparisons with RM2000/v for the latter two predictors

yielded p of 0.23 and 0.60 respectively). For RM2000/v reoffending, OGRS3 and OVP both outpredicted RM2000/v (AUCs 0.78, 0.79 and 0.80; $p < .001$ for both comparisons). For OVP reoffending, OGRS3 and OVP again outpredicted RM2000/v (AUCs 0.80, 0.83 and 0.83; $p < .001$ for both comparisons).

Table 6. Comparison of four-year proven nonsexual violent reoffending outcomes for standardised banded RM2000/v, OGRS and OVP scores, using the RM2000/v total score distribution for all measures

RM2000/v score or equivalent (<i>n</i>)	% with each type of four-year reoffending outcome								
	Homicide & wounding			Violence (RM2000/v definition)			Violence (OVP definition)		
	RMV	OGRS3	OVP	RMV	OGRS3	OVP	RMV	OGRS3	OVP
0 (2,107)	Zero	0.1	Zero	2.0	1.1	1.1	4.5	2.4	2.1
1 (1,595)	0.3	0.1	Zero	5.6	4.2	3.4	10.9	6.8	6.8
2 (1,849)	0.2	0.5	0.4	10.0	9.7	10.2	18.9	17.3	17.3
3 (2,146)	0.8	0.8	1.2	20.2	19.7	20.5	34.8	34.6	34.9
4 (1,762)	1.6	1.9	1.5	33.9	32.3	31.6	54.1	52.5	50.1
5 (1,901)	2.6	3.0	2.9	42.1	43.6	43.4	62.2	64.7	66.4
6 (1,511)	3.2	3.5	3.2	50.3	54.5	54.4	71.2	77.4	78.6
7 (790)	5.6	3.8	4.1	63.9	61.9	60.8	83.0	84.4	83.5
8 (235)	4.7	3.4	5.5	66.4	71.9	76.2	82.6	91.9	92.3
AUC	0.78	0.76	0.77	0.79	0.81	0.81	0.81	0.84	0.85
	(0.75, 0.80)	(0.73, 0.78)	(0.74, 0.79)	(0.78, 0.80)	(0.80, 0.81)	(0.80, 0.81)	(0.81, 0.82)	(0.84, 0.85)	(0.84, 0.85)
					***	***		***	***
All	1.5% (209/13,896)			25.7% (3,569/13,896)			39.1% (5,426/13,896)		

Note. *N* = 13,896. RMV = the total RM2000/v score. The OGRS3 and OVP total scores were both sorted into bins the same size as the RM2000/v score bins, with ties broken randomly. ♦ denotes scores that are significantly worse than RM2000/v using T-test comparisons, while * denotes scores that are significantly better than this tool. ♦ or * *p* < .05

♦♦ or ** *p* < .01

♦♦♦ or *** *p* < .001

Given the failure of the OVP total score to outpredict its static component, logistic regression models were run to compare the predictive contributions of the eleven items and short scales which comprise the OVP 100-point score. Table 7 displays these models. The units of measurements are points on OVP's 100-point scale. Therefore, if all components had equal model parameters (*B*), their predictive contribution for sexual offenders would be identical to that assumed by OVP's scoring system.

Instead, looking first at the homicide/wounding model, sanctions for nonviolent offences, first-time offender status and temper control problems were more highly predictive than the scoring system assumes, whereas offence impact recognition, alcohol misuse and attitude problems were all entirely nonpredictive of this outcome. For the less serious violent outcomes, all four static risk factors were predictive, whilst temper control became less predictive and alcohol misuse became more predictive. The differences between the four static risk factors are partly but not fully explained by the rounding of weights explained in Howard and Dixon (2012a), where the OVP scoring system does not exactly reflect the *B*s observed in the model used to construct it. In summary, nonviolent offending history and first-time offender status appear to be far more predictive among sexual offenders than general offenders, while OVP's constituent dynamic risk factors have either moderate or weak predictive validity among sexual offenders.

Table 7. Logistic regression models examining predictive contributions of OVP's weighted static and dynamic risk factor scores for four-year proven nonsexual violent reoffending outcomes

Risk factor type	Risk factor (maximum points)	Model parameters by four-year proven reoffending outcome					
		Homicide & wounding		Violence (RM2000/v definition)		Violence (OVP definition)	
		<i>B</i>	<i>SE B</i>	<i>B</i>	<i>SE B</i>	<i>B</i>	<i>SE B</i>
Static	Constant	-8.20	0.72	-4.75	0.13	-4.29	0.10
	Sanctions for nonsexual violent offences (25)	0.05	0.02	0.10	0.00	0.13	0.01
	Sanctions for other offences (5)	0.18	0.06	0.16	0.02	0.17	0.02
	Any previous sanctions (5)	0.35	0.15	0.24	0.03	0.20	0.02
Dynamic	Age (20)	0.11	0.01	0.12	0.00	0.14	0.00
	Fails to recognise impact of offence on victims/society (4)	0.00	0.04	-0.05	0.01	-0.04	0.01
	Accommodation problems (4)	0.03	0.05	0.01	0.02	0.02	0.01
	Employability problems (6)	0.00	0.05	0.03	0.01	0.07	0.01
	Alcohol misuse (10)	-0.01	0.02	0.03	0.01	0.06	0.01
	Psychiatric treatment current or pending (4)	0.05	0.06	-0.03	0.02	0.02	0.02
	Temper control problems (6)	0.16	0.04	0.08	0.01	0.06	0.01
	Attitude problems (6)	-0.04	0.06	-0.02	0.02	-0.02	0.02

Note. *N* = 13,896. Gender is also a static risk factor in OVP, but is disregarded in this all-male sample.

Discussion

This study aimed to examine the predictive validity of four violence risk assessment tools with a large sample of convicted sexual offenders in England and Wales. Offenders were followed for four years after release from prison or start of community sentence. Rates of proven nonsexual violent reoffending varied depending on the definition of violence used. However, using the most inclusive definition (OASys Violence Predictor [OVP] -class offences), around 37% of the sample had been proven to have committed a violent reoffence within four years of being in the community. All four of the tools examined (Offender Group Reconviction Scale [OGRS] 3, OVP, RM2000/combined sexual and nonsexual violence scale [RM2000/c] and RM2000/violence scale [RM2000/v]) were able to place offenders into relative risk groups, with those classed as lower risk having lower rates of proven reoffending than those in the higher risk groups. The number of offenders assigned to each risk group varied greatly between tools. Those tools developed on general offender populations (OGRS3 and OVP) and whose very high risk band was designed to include very few offenders, had far fewer high and very high risk offenders than those developed on sexual offender populations (RM2000/c and v).

The predictive accuracy of the tools was fairly equivalent when the proven violent outcome was homicide and wounding; all tools demonstrated good predictive validity, although the RM2000/c was worse than the other tools examined. Given that this tool was designed to assess risk of sexual and violent reconviction it is perhaps not surprising that it did not do as well as those tools that were designed specifically

to predict nonsexual violent recidivism. While predictive validity for RM2000/v- and OVP-class reoffending was also high for all tools, the results of comparisons were very dependent on the way predictive scores were grouped.

When the tools grouped offenders into four risk categories they were less accurate than if a wider range (total scale score) was used, which supports the notion that restriction of range affects the measure of predictive accuracy commonly used; the Area Under Curve statistic. However, in practice it is necessary for actuarial tools to divide offenders into larger groups, to aid the deployment of resources to the groups at higher risk of committing a serious reoffence. As such, when the tools were in banded form (when they split people into low, medium, high or very high risk), the best tool for predicting homicide and wounding reoffences was the RM2000/v, the best tool for predicting RM2000/v-class reoffending was OGRS3, and the best tools for predicting OVP-class reoffending were OGRS3 and OVP. When the bandings of the tools were revised so that a similar number of offenders were housed in each risk group, OGRS3 and OVP had significantly better predictive accuracy than the RM2000/v across two of the three proven violent reoffending outcomes considered.

The OVP static scale was more predictive than the OVP total score, and it was shown that several of OVP's dynamic risk factors did not predict nonsexual violence by sexual offenders. In particular, failure to recognise the impact of the offence on victims or society and holding attitudes supportive of offending were nonpredictive of serious violent reoffending (homicide and wounding) and less predictive than expected of the other violent outcomes. A further study of OVP among a general NOMS population (Howard & Dixon, 2012b) also found that the recognition of offence impact and psychiatric treatment items were no longer predictive, but the present study's results affect a wider range of risk factors. There are a number of

possible explanations for these findings. First, measurement of attitudes in forensic populations is difficult, relying on the drawing of inferences from behaviour, or on self-report, which relies on clearly articulated, consciously accessible thoughts which are liable to presentation bias (Nunes, Firestone & Baldwin, 2007). Examinations of the relationship between offending and scores on measures of criminogenic attitudes in sexual offenders tend to yield either no or small effect sizes, with significant variability across studies (e.g., Hanson & Morton-Bourgon, 2004). Second, the fuzzy definitions of these items in OASys are also likely to have impacted upon the reliability of the scores for these factors. Third, Mann, Hanson and Thornton (2010) note that it is difficult to determine, from offenders' statements, the extent to which criminogenic attitudes are present, as we all do or say things that are contrary to our underlying beliefs; the clinician has a difficult task in disentangling the noncriminogenic, esteem-protecting, post-hoc justifications for offending from statements indicative of underlying criminogenic attitudes. Similarly, a failure to recognise the impact of offending is a concept difficult to define and measure. In addition, it has been argued that this too can be the result of a protective mechanism, helping those who wish to desist from offending to create and maintain a prosocial, 'nonoffender' identity (Burnett & Maruna, 2006; Maruna & LeBel, 2003). Further research is required to properly establish the nature of any relationship between these factors and violent reoffending.

The inconsistent findings for OVP's dynamic risk factors perhaps also explain why OGRS predicts as well as OVP does, given that the OVP factors which predicted best were similar to those in OGRS – previous nonviolent offending, first-time offender status and age, rather than previous violence. These findings (Table 7) and the finding of considerable differences in violent reoffending rates by sexual offence

history (Tables 3 and 4) could spur further research on how to optimise prediction of nonsexual violence among sexual offenders. The differences by sexual offending history correspond closely with findings on offence-specific sexual recidivism on an earlier version of the current sample (Howard et al., 2012), which showed that offenders with a history of indecent image offending – whether or not they had also been sanctioned for other sexual offences – seldom ‘crossed over’ to commit non-indecent image sexual reoffences.

Despite these interesting research implications, the immediate practical implications are limited, as the predictive differences between the predictors are fairly limited. A key practical consideration is that the OVP, unlike the OGRS3 score – which is equally predictive of nonsexual violence for sexual offenders but not general offenders (Howard & Dixon, 2012a) – has the advantage that its predictions include estimated rates of violent reoffending, which can aid practitioners in making risk judgements. Unlike RM2000/v, OVP creates higher-risk groups of a size which is manageable in practice. As such, our findings support contemporary shifts in NOMS practice which are favouring the use of OVP to predict nonsexual violent reoffending by sexual offenders. This promotes consistent practice and reduces the total number of risk scales with which practitioners need to become and remain familiar.

Although OVP outperforms RM2000/v as a prediction tool, the present study suggests that the RM2000/v has good predictive accuracy when predicting a range of violent outcomes. While the RM2000/v scoring system is not as flexible as OVP’s, the results in Table 5 suggest that RM2000/v could be used in a flexible manner. As recidivism rates rise gradually across its 0-8 range, it would be possible for jurisdictions using it to alter its risk banding to one with more practical relevance, without damaging its predictive validity. Another advantage of using the RM2000/v

scale could be that it consists of just three items, which are relatively easy to score, requiring information routinely available to those working in forensic settings (age of the offender and criminal history information). Although this brevity could be seen as an advantage, in practise all sexual offenders assessed by NOMS will have an OASys conducted anyway. Thus, within current UK practise, the fact that OVP has more items would not be much of a practical or resource consideration. This may be more of a practical consideration in other jurisdictions who do not collect the relevant information via OASys-like systems. The several European jurisdictions which have risk assessment systems based on OASys (van Kalmthout & Durnescu, 2008) can, however, be reassured that OVP could be adopted for both general and sexual offenders.

The sample used in this study included offenders whose sexual offences were all committed before the age of 16. Risk Matrix 2000 guidance recommends that the instrument should not be used with these offenders. Their sexual reoffending rates appear in our results to fall into an intermediate range – higher than those observed in populations with no sexual offending history, but lower than other sexual offender groups. Their rates of nonsexual violence were high, even after controlling for risk category. These results suggest that an adjusted risk assessment procedure should be developed for use with this offender subgroup.

Limitations

One of the main limitations of the present study is the relatively short follow-up period used. The RM2000/v has been validated on samples with long follow-up periods, whereas the OVP has been developed and validated on samples with shorter follow-up periods. It could be that the predictive accuracy of RM2000/v fares better

than OVP over longer follow-up periods. Further research would be required when longer follow-ups are available to determine whether this is the case. A further limitation is the use of a proxy RM2000/s which in turn created a proxy RM2000/c score. This may have affected the predictive accuracy of RM2000/c, though the effect should be slight given its limited impact on prediction of sexual offending (Barnett et al., 2010).

A further limitation relates to offences with a sexual element or motivation which are not charged as sexual offences. As Table 1 showed, these can be identified when they are the index offence using OASys data, and form a substantial minority of current sexual offences. However, the MoJPNC data available to study recidivism does not allow identification of such offences. As such, an unknown fraction of our “nonsexual violent” reoffences will actually be offences with a sexual element or motivation. This problem is not solvable in large-scale studies of NOMS data, as case files cannot practically be retrieved in sufficient numbers to allow scrutiny of the many statutory nonsexual violent reoffences, but it should be borne in mind by NOMS practitioners and policymakers who interpret the base rates of statutory nonsexual violent reoffending in this study, and statutory sexual reoffending (which underestimates all sexual reoffending) in studies such as Barnett et al. (2010).

This study has followed a consciously conservative approach, without directional prior hypotheses on the relative merits of different tools or risk factors. That is, no prior data was available to support the superiority of one approach or risk factor over another, and *post hoc* tests were not executed to avoid ‘fishing’ for significant results (Steyerberg, 2010). Subsequent research, on samples which do not overlap with ours, could specify stronger hypotheses and formally test the equalities

of risk factors and perhaps strive to develop new predictive models which include sexual offending type and use dynamic risk factors selectively.

CHAPTER FOUR

IDENTIFYING CHANGES IN THE LIKELIHOOD OF VIOLENT RECIDIVISM: CAUSAL DYNAMIC RISK FACTORS IN THE OASYS VIOLENCE PREDICTOR

While Chapter Two established OVP's risk factors, this chapter examines whether scores on the purportedly dynamic risk factors included in OVP do change as offenders are assessed repeatedly over the course of the community portions of their sentences. It also considers whether the risk factors have causal properties: are changes in risk factor scores prospectively associated with changes in violence risk? Despite the potential benefits to risk assessment theory and practice of determining the value of assessing dynamic risk factors, the few existing studies of this topic have used small samples, and often featured assessment scoring by research assistants. This chapter uses a very large sample of assessments completed in NOMS practice to show that some, though not all, of OVP's 'dynamic' risk factors have causal dynamic properties. It also demonstrates that improvements in prediction do occur as a result of using OVP scores from updated assessments rather than relying on scores from initial risk assessment, though the magnitude of these improvements is quite limited.

The following article was accepted for publication in *Law and Human Behavior*, volume 37, pages 163-174, in 2013.

Law and Human Behavior

Identifying Change in the Likelihood of Violent Recidivism: Causal Dynamic Risk Factors in the OASys Violence Predictor

Philip D. Howard and Louise Dixon

Online First Publication, November 12, 2012. doi: 10.1037/lhb0000012

CITATION

Howard, P. D., & Dixon, L. (2012, November 12). Identifying Change in the Likelihood of Violent Recidivism: Causal Dynamic Risk Factors in the OASys Violence Predictor. *Law and Human Behavior*. Advance online publication. doi: 10.1037/lhb0000012

CHAPTER FIVE

SPECIALIZATION IN AND WITHIN SEXUAL OFFENDING

This chapter examines the extent to which convicted sexual offenders specialize in sexual offending rather than having more varied, 'generalist', criminal careers. It also examines whether specialization in particular types of sexual offending can be observed. By doing so, it provides information on the nature of risk presented by sexual offenders. The results show that indecent images offenders have a less generalist pattern of offending than other sexual offenders, and that specialization in four types of sexual offence does exist to varying degrees. It is demonstrated that prediction of recidivism in each of the four offence types should be enhanced if specialization is taken into account, a finding which has not previously been incorporated into the actuarial risk assessment of sexual offenders.

The following article has been published in Online First form by Sexual Abuse: a Journal of Research and Treatment. The published version should be identical to the version reproduced here with the exception of some changes to header formatting. On page 149, the name Mazerolle has been misspelt as Mazzerolle.

AUTHOR QUERY FORM

Journal title: SAX

Article Number: 486934

Dear Author/Editor,

Greetings, and thank you for publishing with SAGE. Your article has been copyedited, and we have a few queries for you. Please respond to these queries when you submit your changes to the Production Editor.

Thank you for your time and effort.

Please assist us by clarifying the following queries:

No	Query
1	In respect of Query No. 2 in the previous list regarding Level 1 heading, the APA guideline is as follows: "2.05 Introduction Introduce the problem. The body of a manuscript opens with an introduction that presents the specific problem under study and describes the research strategy. Because the introduction is clearly identified by its position in the manuscript, it does not carry a heading labeling it the introduction." In view of the above, the heading "Introduction" needs to be deleted. Hence the request to please consider commencing the text with a Level 1 heading.
2	In the sentence "One new type of sexual offending . . ." please clarify whether the year in "Soothill et al.'s" should be "Soothill et al.'s (2000)."

CHAPTER 6: GENERAL DISCUSSION

This thesis aimed to provide a deeper understanding of some aspects of violent and sexual offending risk, and identify ways in which the prediction of reoffending can be improved. It has accomplished these overarching aims by: the development and validation of an empirically-grounded nonsexual violence risk predictor; demonstrating that this predictor includes causal dynamic risk factors; further validating this predictor for sexual offenders, and increasing the evidence base on specialisation in sexual offending as a precursor to future efforts to similarly improve empirically-grounded prediction in this area.

The thesis's Introduction listed five specific aims. Each of its chapters addressed one aim, the results of which are summarised below together with their implications for forensic practice and future research. Some general conclusions and reflections are then provided.

Aims of the thesis, and relevant results

Aim One: empirical construction of a classification of offences for a predictor of violent reoffending

Chapter 1 executed two analytical steps to produce a classification of violent offences.

The first step was an analysis of OASys data on index offence content. This revealed that characteristics such as actual or threatened violence were features of many types of statutory offence. By definition, such characteristics can be expected to be present in statutory offences

including assaults, offences involving weapons and threats, robbery and contact sexual offences such as rape, and this was indeed demonstrated. These characteristics were also found to be present in statutory public order and criminal damage offences. This result was particularly important for criminal damage offences, which were almost entirely absent from the violence classifications used in existing risk assessment tools and studies of violent recidivism (as discussed in the Introduction).

Chapter 1's second step conducted two recidivism analyses. The first part checked for simple associations between dynamic risk factors and reoffending involving each of the seven potential components of a violence classification, plus an 'all other offences' group for comparison. This found that similar dynamic risk factors were strongly associated with all potential components other than contact sexual offences, and that these were different to the associations for other offences. The second part looked at whether histories of the potential component offences were associated with several different recidivism outcomes, using logistic regression models. The outlying nature of contact sexual offences was confirmed.

By using such empirical methods, this paper therefore developed a classification of violent offences which was quite different to those used elsewhere. The finding that sexual offending should be considered separately from nonsexual violence was not without precedent – as noted in the paper, Hanson and Morton-Bourgon (2009) found through meta-analysis that such separate treatment is associated with better prediction of sexual recidivism – but it is useful to have established it directly rather than through comparison of different studies. The breadth of the violence classification, even without sexual offences, makes these results stand out from the classifications used in other research. The results in Table IV, showing the

predictive contributions made by previous sanctions for ‘borderline’ offences such as threats/harassment, public order and criminal damage, sets the scene for the strong predictive performance demonstrated by the OASys Violence Predictor (OVP) in Chapter 2.

Aim Two: construction and validation of the OASys Violence Predictor (OVP)

Chapter 2 used this classification of violent offences as the recidivism outcome for a new actuarial instrument, OVP, which included both static and dynamic risk factors. A sample of 15,918 assessments completed between 2002 and 2004 was used as a construction dataset, and a further sample of 49,346 assessments from 2004-05 as a validation dataset. Static risk factors – age, sex, and violent and general criminal history – and items from OASys's ten domains of dynamic risk were available for selection. The selected model comprised five static and seven dynamic risk factors, with age and violent criminal history attracting the highest weights. In response to feedback from probation officers – the eventual users of any mainstream National Offender Management Service risk assessment instrument - attempts were made to round risk factor weights, and it proved possible to round these considerably without reducing predictive validity appreciably. Comparisons with other risk assessment tools which could be used in NOMS confirmed that OVP was both clinically and statistically significantly more predictive of very serious (homicide and wounding), serious (homicide and assault) and broader violent reoffending.

In recent years, some researchers have proposed that the upper limits of predictive validity have been reached (Campbell, French, & Gendreau, 2007; Yang, Wong, & Coid, 2010). It is likely that a ceiling to predictive accuracy does exist, given that the stochastic nature of

human behaviour and the vagaries of the criminal justice system obscure the relationship between risk factors and, successively, actual and measured recidivism. However, OVP has improved risk prediction through a number of technical efficiencies, including the use of a violence measure (from Chapter One) which enables best use to be made of information on previous offending and, in contrast to other risk assessment systems such as LSI-R, the separation of alcohol and drug misuse. It is probable that further improvements could be made, and some suggestions are discussed below.

This chapter also presented detailed but clear information on the trade-off between false negative and false positive results, using the specificity and sensitivity measures which lie underneath the Area Under Curve predictive validity measure. Presenting information on predictive validity in a manner which can be clearly understood outside the research community is vital if practitioners are to make correct use of risk scales (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Hanson & Howard, 2010), and I continue to seek opportunities to practically illustrate the benefits of improving risk prediction.

OVP was implemented across NOMS in August 2009, and has been a feature of approximately 2,400,000 assessments of 600,000 different offenders, as of February 2013.

Aim Three: validate OVP for offenders with a history of sexual offending

Chapter 3 considered whether it is appropriate to use OVP as a predictor of nonsexual violence among offenders with a history of sexual offending. It has been argued (David Thornton, personal communication, October 2010) that prediction of nonsexual violence is

especially important for sexual offenders because the statutory charge for which an offender is convicted can sometimes obscure the sexual content of an offence. It does not however follow that a different actuarial tool is necessary. Chapter 3 therefore compared the predictive validity of OVP and the Offender Group Reconviction Scale version 3 (OGRS3) with the Violence and Combined scales of Risk Matrix 2000 (RM2000/v and /c), for the three violent outcomes used in Chapter 2. The task was complicated by the very different distributions of the tools: Thornton's risk aversion led to RM2000 including many offenders in its higher risk bands, whereas a focus on allocating NOMS's limited resources led to OVP and OGRS3 having small higher-risk bands. The Chapter controlled for this issue by using the underlying risk scores to create similar sized risk bands, and found that OVP was somewhat superior.

Two further aspects of nonsexual prediction were explored. Firstly, as analysis revealed that scores on the static aspects of OVP were at least as predictive as the OVP total score, logistic regression models were used to examine the predictive performance of each of OVP's component risk factors. This revealed that only some dynamic risk factors were predictive among this subpopulation, with temper control being the most important across the three outcomes. Secondly, the effect of specialization within sexual offending (see also Chapter 5) was considered by comparing sexual and nonsexual violent reoffending rates for offenders with different sexual offence histories. Those with histories of indecent image offences had lower rates of nonsexual violent recidivism, those who had not offended sexually since their sixteenth birthday (which RM2000 guidance advises should not be actuarially assessed) had lower but not negligible rates of sexual recidivism, and those with prior rather than current sexual offences had higher rates of nonsexual violent recidivism.

The results of this paper have considerable practical value for NOMS: they show that it is not necessary to maintain separate nonsexual violence prediction procedures for offenders with a history of sexual offending. Given that sexual offenders will receive OASys and therefore be scored on OVP, NOMS has been enabled to streamline risk assessment and communication processes by not using RM2000/v or /c. The RM2000/c scale had already been removed after Barnett, Wakeling, and Howard (2010) found it unhelpful in a study of RM2000's three subscales, and NOMS is now considering whether there is any future need for RM2000/V. This paper does though also indicate that OVP's dynamic risk factors have limited relevance among sexual offenders, raising the question of whether there may be other dynamic risk factors for nonsexual violence by sexual offenders (and, if not, why not?).

Aim Four: measure changes in OVP scores and subscales, and determine whether these changes are associated with changes in the likelihood of violent recidivism

Chapter 4's literature review found that existing evidence on causal dynamic risk factors was limited and covered small samples, but generally supported the hypothesis that social/personal risk factors can change and that revising risk scores improves risk prediction. The chapter then introduced an extremely large sample, which was used to trace 196,493 cases from initial assessment to reassessment and/or recidivism for violent (OVP's classification) and homicide/wounding recidivism outcomes. For each assessment, the total OVP score and its dynamic risk factor components were separated into initial assessment scores and changes from initial to current assessment; this split allowed the initial and change elements to be examined separately in survival models. Basic checks showed that only a slight majority of all nonrecidivist offenders were reassessed within the timescales recommended in NOMS

guidance, and that many reassessments claimed no change in any of the 62 dynamic items in OASys at this time. In short, this study was completed in real world conditions which severely tested the practical relevance of the theoretically appealing causal dynamic risk factor construct: many potential changes in risk factor scores were missed due to assessors' failure to conduct any reassessment, and other changes must surely have been unrecorded by reassessments which claimed no change across a panoply of risk factors.

A major result concerned change in total OVP scores: even though reoffenders had higher scores at initial assessment than nonreoffenders (and therefore reoffenders had less 'room' to increase their scores at reassessment), reoffenders experienced net increases in OVP dynamic scores at reassessments conducted prior to their reoffending, while nonreoffenders' net scores fell further as time went by. This is a useful high-level result for academic readers, but it is perhaps most important because it can be easily communicated to nonspecialists such as NOMS policymakers, in the style of, "Offenders who eventually reoffended tended to have increases in their dynamic risk scores when they were reassessed, while offenders who did not go on to reoffend tended to see their scores decrease".

This simple result was confirmed by more technically sophisticated analyses of survival data. Measures of the OVP score's predictive validity were better for sequences of assessments (i.e. allowing offenders' scores to change over time) than for single-point assessment (i.e. retaining the initial score and disregarding reassessments). Cox regression models demonstrated that the initial score and the change in score had similar hazard ratios for both recidivism outcomes, thus indicating that their sum – the OVP score at reassessment – is a 'fair' score which can be used without any reweighting or other adjustment.

Similar packages of methods were then used to investigate the nature of each dynamic risk factor. There was a range of results: alcohol misuse scores often changed and their score changes were highly predictive. Temper control changes were less frequent but highly predictive, especially of homicide/wounding. Employment, accommodation and attitudes changes were frequent and moderately predictive. However, the remaining two risk factors – recognising the impact of offending, and psychiatric treatment – did not demonstrate the properties of causal dynamic risk factors, changing rarely and with inconsistent predictive value when they did change. Future developments in violence risk prediction could benefit both from these specific findings and from the development of a methodological framework which encourages the selection of causal dynamic risk factors.

This chapter's results provide several other worthwhile insights. By demonstrating more conclusively than any previous research that causal dynamic risk factors do exist and can be used in practice, they bolster the 'needs' element of the Risk Needs Responsivity approach to offender assessment and treatment. Given that changes in certain risk factors are proven to predict reoffending, evaluation protocols could be developed which exploit score changes as intermediate outcome measures. As pressures on public finances produce an environment where probation resources are subject to severe pressure, this chapter provides a timely quantification of a major benefit (though not yet the cost-benefit) of repeating OASys assessments.

Aim Five: understand the extent of specialisation in particular types of sexual offending and in sexual rather than general offending

Chapter 5 reviewed the existing evidence on sexual offence specialisation. It found little evidence of specialisation in sexual offending, and some evidence of specialisation within sexual offending. It then proceeds to examine both types of specialisation in a sample of 14,804 NOMS offenders; this sample was a subset of the type of general OASys-assessed sample used in previous chapters, using only those with a known history of sexual offending. The presence of specialisation was tested both retrospectively (examining criminal histories) and prospectively (examining patterns and predictors of reoffending).

The most persistent pattern in the paper was a difference between offenders convicted of indecent images offences and those convicted of other sexual offences. Indecent image offenders were much less likely to have previous convictions for nonsexual, nonviolent offences and for nonsexual violent offences. They were also far less likely to recidivate in any of these categories (Chapter 3 gave further detail on this by examining three categories of nonsexual violent recidivism). Across all four categories of sexual offences – the other three being contact adult, contact child and paraphilia – there was considerable specialisation in the criminal histories of those with two past sanctions for sexual offending. Predictive models of sexual recidivism also revealed offenders' tendencies to continue in the same type of sexual offending. Some other predictors, such as stranger and male victims, and not being a first-time offender, were also differentially associated with different risk factors. The final set of results confirm that predicting all sexual recidivism with a single model leads to suboptimal predictive validity, with offence-specific models offering nontrivial improvements.

While specialisation within sexual offending had been observed in the literature explored in

this Chapter's literature review, many of the methods used in the data analysis had not previously been applied to sexual offence specialisation. This allowed strong emphases to be placed upon the links between past and future specialisation and the implications for actuarial prediction. Understanding specialization within sexual offending is important in risk prediction, enabling identification of offenders likely to commit the most dangerous reoffences and therefore enhancing public protection. A fuller understanding of specialization may also improve understanding of treatment targets, if dynamic risk factors prove to differ between sexual offenders with different offending patterns. Further analysis of this type of data might add further to knowledge by, for example, applying the Forward Specialization Coefficient (Farrington, Snyder, & Finnegan, 1988) to the criminal careers of those convicted of different types of sexual offence.

Implications and limitations of findings and future directions for research

Implications of the thesis findings

Some of the implications of this thesis's findings have been set out above, especially the practical implications for NOMS. In this section, I mention further implications for research into the prediction of reoffending, including how it is informed by an understanding of offence specialisation.

As the introduction to Chapter 2 explains, and the summary above reiterates, some researchers believe that the success of risk prediction methods has reached a “glass ceiling”. The research in this thesis has demonstrated three objections to this pessimistic statement. Firstly, as

Chapter 2 shows, a broad range of dynamic risk factors can, modestly, improve predictive validity. Secondly, allowing reassessment of these risk factors over time also improves validity. Finally, understanding how offenders specialise in particular offence types makes it possible to better target risk predictors to create a cohesive set of offences which are committed by similar offenders in the past and future.

The relatively strong predictive validity achieved by OVP demonstrates that these steps can raise the glass ceiling to some extent. In a similar manner, but beyond the scope of the five chapters of this PhD, I have taken forward the findings on sexual offence specialisation in Chapter 5 to create the OASys Sexual reoffending Predictor (OSP; Howard & Barnett, in press), which takes contact sexual offences as its recidivism outcome and as such differentially scores histories of contact adult, contact child, paraphilia and indecent image offending. The use of a single outcome was felt to be a necessary compromise in view of the practical difficulties associated with a proliferation of risk scores, identified in Chapter 5's discussion.

A further step, under development, is the tightening of the focus of violent prediction onto the most serious offences of homicide and wounding; preliminary findings suggest (as might be expected from the model of this outcome in Chapter 1) that an augmentation of OVP with additional factors related to criminal histories of serious violence may be successful.

Limitations of this research

Available data.

One limitation of the research in this thesis has been its reliance on a single risk assessment instrument as a source of dynamic risk factor information. OASys was designed between 1999 and 2001, and has changed little since then. (Its most important change was an update in 2009 which introduced OVP and the OASys General reoffending Predictor, but also reduced the total length of the assessment for cost-saving reasons.) As has been detailed at various points earlier in the thesis, the validity and reliability of OASys is either unknown or suboptimal in several respects: Chapter 4 mentioned deficiencies in the frequency and quality of reviews, Moore (2009) found that some of the dynamic risk factor scales had greater construct validity than others, there have been no studies of test-retest reliability, and the only studies of inter-rater reliability (Howard, Clark & Garnham, 2006; Morton, 2009) contained serious flaws and cannot be viewed as authoritative. As Chapter 4 therefore emphasises, this thesis reports the achievements which can be made using data obtained from “realistic correctional conditions” (pg. 134). It may be that greater progress could be made if more resources were dedicated to both researching the reliability and validity of OASys and ensuring that assessor and supervisor caseloads were light enough to allow assessments to be completed to a consistently high standard.

Several areas of potential improvement to OASys can be identified from the perspective of estimating, managing and treating violence risk. The additional risk factors which might be considered for inclusion are listed in the next section.

More frequent assessment would also make it more likely that acute dynamic risk factors would demonstrate value in risk prediction and management, as they are unlikely to

demonstrate predictive validity if only assessed every four months as prescribed until recently under NOMS National Standards. Indeed, one interesting additional analysis considered for inclusion in Chapter 4 – which was edited for length - demonstrated the 'decay' in the incremental predictive value of dynamic risk factors as time passed since the most recent assessment. Unfortunately, budgetary pressures in NOMS make it likely that OASys updates will occur less frequently rather than more frequently in the future. At the time of writing, it is unclear what risk assessment processes will be required to be followed by providers of probation services under contract, as part of the recently announced Transforming Rehabilitation programme (Ministry of Justice, 2013).

PNC data also has its limitations. It is not possible to study recidivism involving domestic violence, as this is not a separate criminal offence and therefore cannot be distinguished from other types of violence. Historical record transfers, from the pre-1995 microfiche system to the modern-day PNC, and data weeding exercises may result in incomplete criminal records being available. (As a consolation, it is worth noting that researchers and operational users are at least accessing identical data, so flaws in the datasets used to construct and validate predictors of recidivism are identical to those in the criminal history data used in correctional practice). The PNC data extracts also make no reference to death, serious ill-health and emigration, so we must accept that not all of the apparent 'nonreoffenders' are individuals who are alive, well and at risk of being convicted of offences in England and Wales.

Study length.

The efficiency of data analysis increases through the thesis, as the author's technical

knowledge developed. The majority of studies used fixed reoffending followup periods, though Chapter 2 divided this into shorter periods to obtain an ordinal regression model. Survival analysis, as employed in Chapter 4, makes fullest use of the data, and should be used in those future studies which are focused upon exploring the properties of dynamic risk factors rather than constructing or validating predictions for fixed followup periods.

OVP has not yet been validated over long follow-up periods, nor have the patterns of sexual specialisation in Chapter 5 been confirmed over a period of several years. Peer reviewers on Chapters 3 and 5 have commented upon the importance of long follow-up periods for sexual recidivism, due to its low base rate and the potential for delayed reoffending processes as offenders “go to ground” and gradually build links with potential new victims. The low base rate is of some concern in that it leads to predictions having confidence intervals which are broad in relative terms, given the finite size of sexual offender samples even given my access to samples for the whole of England and Wales, and further validation work should continue in order to improve the precision of risk estimates for all follow-up periods.

There appears to be some, but limited, value to the criticism around reoffending processes. My paper for the Ministry of Justice on hazards of reoffending (Howard, 2011) showed that the hazard of sexual reoffending does fall over time, but falls more slowly than the hazards of most other reoffence types. The challenge I have posed both in responses to peer reviewers and in meetings at the National Offender Management Service, which has never been answered, is for an explanation of any mechanism which might prevent short-term results (e.g. the predictive validity of a risk score, or a specialisation pattern) being generalisable to the long term. That is, what risk factors might cause offenders to behave in one way in the

first few years of follow-up but in another way in subsequent years? One possible argument might come from the continuing debate on offence specialisation (see the Introduction): if offenders specialise in particular offences more in the short term than the longer term, then the predictive value of offence-specific criminal histories would diminish in the long term and therefore actuarial risk predictors would become less effective. (All actuarial risk predictors for violent and/or sexual reoffending include offence-specific criminal history as a major risk factor.) Long follow-ups should therefore be generated to provide conclusive evidence on this topic.

One further note on this topic is that most chapters use a 12 month ‘buffer’ period to allow convictions to be incurred and data to be entered onto the PNC. (For example, in Chapter 2’s validation sample, a 2-year reoffending period follows the most recently assessed offenders from September 2005 to September 2007; PNC data extraction in November 2008 allowed slightly over one year for even the very last of these reoffences to become PNC-recorded convictions.) For Chapter 3, I was persuaded by Ministry of Justice statisticians, who had reviewed successive PNC data extracts, that a 6 month ‘buffer’ was acceptable, as the reoffending status of very few offenders was affected whereas the number of offenders who could be included would increase substantially. Ideally, a formal study would set out the costs and benefits of varying the ‘buffer’ period, estimating the cost (i.e., undetected reoffences) separately for the most serious offences as these may take longest to be brought to justice.

Statistical methods.

The samples of all five chapters, and most notably Chapter 4, allow individual offenders to be

included in the sample with multiple observations. I am grateful to the external reviewer of this thesis for proposing a methodological improvement which could be made: the use of variance inflation factors to control for the impact of this repetition on measurements of statistical error. As these factors were not used, the standard errors of parameter estimates in this thesis's regression models are somewhat too narrow, and p values for statistical significance should therefore have been somewhat larger.

New sexual offending by those with no known history of this offence type.

Chapters 3 and 5 both refer to sexual recidivism by those offenders with a known history of sexual offending. This is 'standard practice' for research on this topic. However, this approach ignores another type of 'reoffending': first-time sexual offending by those with other criminal histories. In unpublished data on two-year community follow-ups starting in 2008/09, I recently found that only 100 of 314 contact sexual offenders had a history of (proven) sexual offending. While the base rate is much higher among this with a known history (who make up less than one-tenth of the sample), this topic deserves further investigation if NOMS is to be able to properly compare "total dangerousness" (i.e., violent and sexual offending risk) between all offenders on its caseload.

The only published study on first-time sexual offending is by Duwe (2012), which achieved considerable predictive validity among a sample of discharged Minnesotan prisoners. Duwe's model includes some items with questionable face validity (e.g., "robbery convictions [while aged] under 21"), but essentially appears to reflect general antisociality and perhaps personal trauma. Evidently, further investigation of this topic may have considerable value both in

terms of understanding transitions into sexual offending and in more accurately describing offenders' aggregate risk across a range of potential harmful behaviours.

The validity of actuarial risk assessment.

A more fundamental limitation to the application of this thesis's findings could be that raised by David Cooke, Stephen Hart and Christine Michie (*Cooke et al.*) in a recent series of papers (Hart, Michie, & Cooke, 2007; Cooke & Michie, 2010; Hart & Cooke, 2013). Fundamentally, these papers present two arguments. Firstly, the confidence intervals around actuarial risk predictions ought to be quoted but are not and indeed are often incalculable due to the methods used to create them. Secondly, that actuarial risk predictions also have “precision intervals”: essentially, that the predicted recidivism probability for an individual ought to be expressed not just as a point estimate but also with an interval (for which they provide formulae) within which that individual's predicted probability might actually fall after allowing for latent variation between individuals in the group. As the 2010 and 2013 papers complained that early critics (e.g., Mossman & Sellke, 2007) misunderstood these arguments, it may be useful to provide a hypothetical example. An actuarial instrument might provide a predicted reconviction rate of 10% for the group of individuals in a certain risk band. Following Cooke et al.'s recommendations, a confidence interval of 8% to 12% for this group's predicted rate is calculated and presented, to recognise that the limited sample size used to construct the instrument leave uncertainty around the estimated recidivism rate for the risk band. They also recommend that a much wider precision interval is calculated for the individual – this might have a range of 1% to 60%. This is because while the central estimate of the mean risk of the group is 10%, individuals within the group really have very different,

unobservable, actuarial risks which merely have a mean of 10%.

The first argument has some merit, especially for instruments created using a relatively small sample, although it is not evident that real-world decisionmakers (e.g. judges and magistrates receiving Pre Sentence Reports) would welcome the additional information provided by a confidence interval. Its relevance for this thesis is extremely limited, as while the confidence intervals surrounding any given OVP estimate could be calculated, it would be extremely narrow given the vast construction and validation samples utilised.

The mathematical reasoning underpinning the second argument has been questioned (e.g., Harris, Rice & Quinsey, 2007) but, whether technically correct or not, this argument essentially illuminates an underlying truism: an actuarial risk prediction for an individual offender is an estimate, and ought not be given a label of infallibility by either the proponents or critics of actuarial methods. The inherent fallibility of any and all risk predictions ought to be obvious to all observers, and made obvious by the tool's developers to remove any doubt. In the case of OVP, it is clearly stated – for example, through the communication of Area Under Curve statistics and classification metrics such as false positive and false negative rates – that actuarial instruments do not perfectly distinguish potential reoffenders from potential nonreoffenders. It is also readily apparent that no actuarial instrument is based upon a truly comprehensive evaluation of the offender's life history and present personality, behaviour and life circumstances. Therefore, the risk estimates are just that: estimates, overtly based on how a limited set of data items have been associated with reoffending among a sample of past offenders. Proponents of actuarial assessment, including myself, claim not that these methods are perfect but rather that they are the best option available.

I was the co-author of one of the published responses to Cooke et al.'s first two papers (Hanson & Howard, 2010). Our response recommends humility in the communication of risk estimates, including the acknowledgement that they are produced on the basis of group data, and the specification of the information upon which any estimate is based. We also note that structured clinical risk ratings have the same limitations, which are hidden but not negated by the lack of a numerical estimate.

Karl Hanson and I would agree with Cooke et al. that the presence of unobserved and often unobservable risk factors means that no reasonable person will believe that every individual with an actuarial risk estimate of 10% has a true probability of that risk event of exactly 10%. The discussion points made by Hart and Cooke (2013) suggest that some North American jurisdictions are being misinformed by unreasonable or irresponsible users of actuarial methods who are presenting these predictions as infallible fact in high-stakes situations such as court determinations of Sexually Violent Predator status. We applaud Hart and Cooke's efforts to undermine such practice, though we think they go too far. We maintain that correctional agencies must do their best to assess and manage risk using the information available for each case, and they must conduct these efforts using limited resources which prevent the application of structured professional judgement to the majority of cases. In this situation, encouraging the use of valid but fallible actuarial risk estimates – in the knowledge of how and why they are imperfect, and communicating caveats and limitations where practicable – will lead to better outcomes than assuming that no distinction can be made between lower- and high-risk offenders.

Future directions of research aimed at improving actuarial risk prediction

Beyond the exploration of longer follow-up periods and more frequent risk assessment proposed above, at least three sources of potential improvement in predictive validity can be identified. These are the use of complex statistical methods to produce the prediction algorithm, developing and integrating understanding of how hazards (short-term risk) of reoffending change over time, and considering additional risk factors. (For an earlier perspective on improving actuarial risk prediction, see Kroner et al., 2007.) The value of combining actuarial and clinical risk estimation methods should also be considered, and outcome measures other than binary reoffending ought to be studied.

Complex statistical methods.

Several recent papers have experimented with the use of complex methods to predict violent recidivism. At least two early efforts appear to have ran aground due to a failure to appreciate shrinkage in predictive validity when new samples were used. The Classification of Violence Risk instrument (COVR: in Monahan et al., 2005, its authors cheerfully report both their initial success and the later disappointment) used multiple classification trees, while Dow, Jones and Mott (2005) created an application which used nonlinear methods to estimate offenders' outcomes on the basis of their similarity to a small number of similar offenders with known outcomes. This method resulted in extraordinarily high predictive validity among 'exemplar' cases, but its extreme vulnerability to shrinkage is a likely explanation of the lack of any further published studies on the application.

The methods used in more recent studies seem more robust, but do not yet provide compelling evidence in favour of complex methods. Berk, Sherman, Barnes, Kurtz and Ahlman (2008) used machine learning techniques to estimate actual or attempted homicide in Philadelphia. The results from their validation sample appeared impressive, but cannot be evaluated properly as the authors provided no credible comparison with simpler methods. (Their preferred method flags a certain number of cases as 'high risk'. They then claimed that logistic regression failed to identify high risk offenders as only 2 of 30,000 logistic regression estimates were above 0.5. It is of course no more sensible to use 0.5 as a probabilistic threshold in homicide prediction than it would be in the detection of large Earthbound asteroids; a fairer comparison would involve the creation of higher-risk cohorts of equivalent size, as in Chapters 2 and 3 here.) A more pessimistic view came from Liu, Yang, Ramsay, Li and Coid (2011), who found that neural networks predicted recidivism by English and Welsh prisoners only fractionally better than classification trees and logistic regression.

Making efficient use of data may be just as productive as using a complex method, though the two are not mutually exclusive. An example of efficient data use has been provided by Duwe (2012) while investigating the transition from nonsexual to sexual offending: this paper avoided the need to set cases aside for validation, by using bootstrapping to estimate model shrinkage. In my future research, I hope to make use of efficiencies such as this; I often refer to Steyerberg (2009) for comprehensive guidance on many aspects of model fitting.

Hazards of reoffending and offence-free time.

Actuarial predictors of reoffending, including OVP, take no account of indications that the

offender may have desisted from offending. Predictors are calibrated from samples of offenders who have just been released from prison or commenced community sentences. Evidently, offenders' risk of reoffending must diminish at some point: the courts and correctional services are not clogged up with elderly offenders who have reoffended after going straight for decades. The practical question which follows is thus whether the risk of reoffending falls quickly enough to affect risk management processes over the course of community correctional supervision.

I investigated this question in a paper published by the Ministry of Justice (Howard, 2011). This traced the hazard (short-term conditional probability) of reoffending over the first four years following community sentence or discharge from custody. The hazard fell over time for overall reoffending and for every offence type: that is, an offender's probability of reoffending in period $X+1$, if they had not reoffended by the end of period X , was lower than their probability in period X .

Despite its absence from risk predictors, this question has not been absent from the wider literature. Harris and Hanson (2012) followed up a large sample of Canadian sexual offenders, and concluded that those with a Medium risk score on Static-99 presented no more sexual recidivism risk than nonsexual offenders once 14 years had passed without a sexual offence, and High risk offenders for 17 years. (Low risk offenders always had a very low sexual recidivism risk.) Similar issues have been investigated in the quantitative criminology literature (e.g. Bushway, Nieuwbeerta, & Blokland, 2011).

Information on time passed since last offence could therefore be integrated into risk predictors

by treating this *offence-free time* as a type of risk factor (or, rather, protective factor). Since the work presented in this thesis, I have integrated offence-free time into new generations of OVP and OGP. OVP version 2 (OVP2; Howard, in press) therefore splits its 100-point scales three ways, adding offence-free time to recalibrated versions of the static and dynamic subscales. The method used here was simply to fit polynomial terms for offence-free time and convert the parameter estimates into a month-by-month schedule of points scores. It is possible that this approach may also be used for sexual recidivism in the future, though the relatively slow decline in this hazard (Howard, 2011; Harris & Hanson, 2012) plus the low base rate makes accurate estimation more difficult without a longer followup period. Further statistical modelling possibilities, such as the use of parametric or semi-parametric survival analysis rather than the more familiar Cox regression, are described by Singer and Willett (2003).

Risk factors which have not yet been assessed in OASys.

The thesis Introduction gave an overview of various risk factors for violent offending. Many of these, including age, history of antisocial behaviour (i.e., criminal history, especially for violence), alcohol misuse, antisocial attitudes and negative affectivity (i.e., in OVP, temper control), are included in OVP. Impulsivity is another such factor, which proved to be a significant predictor in the analysis for OVP version 2. Other factors (co-offending, neighbourhood-related factors) were considered in initial analyses of OASys data but had little predictive value. However, several of the risk factors for violent reoffending, and most of the risk factors for sexual reoffending, are not currently assessed or are assessed inadequately.

The putative individual risk factor for violent offending which is most clearly not assessed at all is nutrition (Gesch et al., 2002). Antisocial associates were not a significant risk factor in OVP, but the OASys questions on this topic have no clear relationship with actual or potential violence, and the lack of a question on gang membership is viewed as an omission by many stakeholders in NOMS (though Smithson, Ralphs & Williams, 2013, raise convincing and disturbing concerns about the conflation of ethnicity and gang involvement by criminal justice system agents in contact with non-white British youth). A question about severe head injury in the emotional well-being section of OASys was not significantly associated with violence in preliminary analyses in the construction of OVP, but its frequency was far below that identified by Williams (2012). Further research might therefore check whether OASys assessments are sufficiently thorough to detect the thoroughness of traumatic brain injury. The emotional well-being section also does not focus specifically on psychosis or treatment alliance and adherence, identified as promising risk factors (see the Introduction) by Douglas and Skeem (2005).

OVP successfully includes some general questions about attitudes, but it is possible that more could be done, as attitudes towards the use of violence are not probed. I am also interested in hostile attribution as a risk factor, because it is plausible that the offender's propensity to negatively interpret others' words and actions will relate to the likelihood that a risky situation in which they are involved will escalate to actual violence, as stated in the cognitive reassociation model (Berkowitz, 1993). Another part of the cognitive reassociation model, the pathway from negative affect to bodily arousal to violent reaction, may be well represented by OASys's temper control question, which is an important component of OVP.

In turn, the likelihood of offenders finding themselves in such situations should be considered: OASys's questions on alcohol consumption and relationship quality will relate to this. (Relationship quality is a scored risk factor in OVP2 though not OVP; problematic leisure circumstances and antisocial associates are therefore the only two of the eight general offending risk factors of Andrews and Bonta, 2003, which have not yet been associated with violent recidivism in NOMS.) There have also been interesting efforts in recent years to consider the circumstances and constituent behaviours of violent and sexual offences (e.g., Lehmann, Goodwill, Gallasch-Nemitz, Biedermann, & Dahle, 2013) and to assess situational factors for violence (PRISM; Cooke & Johnstone, 2010). Despite this, from the perspective of those developing assessment tools such as OASys which are to be used on a large number of offenders, there is not yet a sufficiently quick and simple way of measuring the offender's propensity to "get themselves into" situations in which they may then act violently. Work I am currently undertaking suggests that this could be especially important for the prediction of very serious violence, as some items which indirectly suggest immersion in a wholly criminal lifestyle (type of current offence, pro-criminal attitudes) may have predictive value despite their weak validity as a measure of the criterion of interest.

The personal and social risk factors used in OASys and most other risk assessment instruments tend to be traceable back to the psychology of criminal careers (Andrews & Bonta, 2003). One exception is the Psychopathy Checklist – Revised (PCL-R; Hare, 2004), which focuses on a personality disorder. While some questions similar to those of PCL-R are included in OASys, it does not systematically assess psychopathic personality, nor any other personality disorders, and most of its questions about psychology and psychiatry are vague around the type of disorder involved. This is perhaps unavoidable, given the constraints on the

length of the assessment and the limited professional skills of OASys assessors, who would be able to comprehend medical reports but not assess mental health status themselves. One possible source of improvement in this area might be the use of self-assessment personality questionnaires, as these reveal interesting correlations between personality factors from these questionnaires and aggression and antisocial behaviour (Miller, Lynam, & Leukeful, 2003; Miller, Zeichner, & Wilson, 2010; Rolison, Hanoch, & Gummerum, 2013). Whether offenders would complete these questionnaires honestly in potentially high-stakes situations, and if these factors would demonstrate incremental predictive validity, has not yet been tested. At very least, a pilot study (where nothing were at stake for the offenders) might illustrate correlations between personality, OASys items (including those in OVP) and reoffending outcomes, which might then be used in more conventional OASys development.

There is less to say about improving the use of dynamic risk factors in actuarial sexual offender assessment, for the (perhaps counterintuitive) reason that the field is less well developed. Comparing the excellent overview of these dynamic risk factors by Mann, Hanson and Thornton (2010) with the contents of the risk assessment instruments available, it is apparent that dynamic factors are only included in structured clinical tools (e.g., SVR-20: Boer, 1997) which, in turn, are relatively neglectful of two important static risk factors, age and general criminality. Moreover, no single tool includes all or most of the risk factors which Mann et al. regard as proven or highly promising. NOMS is currently developing a tool for the assessment of stable and acute dynamic risk factors; this will benefit from such evidence, and it is possible that it could eventually be used alongside an actuarial tool like OSP. From my work within NOMS, it is clear that considerable further progress will be necessary in order to produce a user-friendly tool that can be reliably completed by operational staff within

practical time constraints.

The introduction of this thesis also includes a short section on neurobiological correlates of violence. It is possible (Beech, 2008) that risk assessment in the future will be greatly enhanced by understanding of individual offenders' cognitive deficits and organic brain disorders, which could be assessed with the assistance of brain imaging methods. Given that affordable methods are unlikely to be available for some years, and this thesis focuses on routine risk assessment practice, I will leave this topic to be addressed properly by other researchers.

Finally, ethnicity can be quite readily included in predictive models using the existing data, but has not been. This is a deliberate omission: given concerns that the criminal justice system may involve bias (whether conscious or unconscious) against individuals from ethnic minorities, the inclusion of ethnicity in a model could further institutionalise unequal treatment, by treating increases in recidivism probability associated with institutional practices as if they were instead characteristics of the offender. Whereas the predictive roles of age and gender can be explained without controversy through their associations with physical ability to cause injury and patterns of socialising which lead to exposure to potentially violent situations, similar explanations for particular ethnic groups must be approached with considerable scepticism.

As a general rule, it must be acknowledged that the addition of further risk factors will not be a panacea. There must be some ceiling to the validity of predictors of harmful recidivism, due to both the stochastic nature of human behaviour involving rare events (a potential reoffence

may or may not happen due to the precise location of the offender and potential victim, acute dynamic risk factors (e.g., mood, intoxication) and perhaps the behaviour of the potential victim and/or third parties) and the imperfect measurement of outcomes (not all reoffences result in criminal sanction, and some small fraction of those who did not commit any reoffence will be incorrectly criminally sanctioned). Moreover, some offenders will always be 'medium risk': they have problems on some major risk factors but not others, and therefore their anticipated frequency of committing new harmful offences will be such that they may or may not do so within a followup period of one or several years. (And, eventually, they will grow older and move to a more stable cognitive state and life circumstances, and that anticipated frequency will diminish so that they may never reoffend.)

Nevertheless, some statements can be made about the value of new risk markers (i.e., measurements of risk factors). These have been quantified through simulation studies in a recent article (Austin & Steyerberg, 2013), which happily provides solid evidence for several insights and empirical observations which I had independently made over the course of studying for this thesis. New risk markers are more valuable (i.e., increase predictive validity more) when added to a model which has relatively weak predictive validity: that is, the ceiling referred to above will limit their utility when we already know a good deal about the offender. New risk markers are more valuable when they have a prevalence [assuming that they are binary] close to 0.5, and these markers are also more valuable when they are poorly correlated with predictions made using the existing model (i.e., with the risk markers in the existing model). All three of these properties can be understood when it is recalled that the AUC predictive validity measure essentially summarises the model's ability to rank offenders in such a way that reoffenders have a higher rank than nonreoffenders. Therefore, a change to

the model is most productive when it “shakes up” the rankings – so a prevalence near to 0.5 is good as it means that plenty of offenders might be moved around the rankings. The new marker will also make more difference if the existing model is weak because many reoffenders need ‘help’ to get to a higher ranking; if the marker has a low correlation with the existing model, then offenders who have been misclassified on the existing model are not especially likely to also be misclassified by the new marker. That is, when these properties are in place, there is more scope for the new marker to make a difference. It should also be acknowledged that adding to the number of risk markers collected within a single risk domain can still help to “reduce sampling error and produce more reliable results” (Babchishin, Hanson & Helmus, 2012, p.455) and thus improve prediction (Babchishin et al. examined the effect of combining correlated actuarial risk scales for sexual offending).

We can apply these insights to pose a series of questions which should be asked when considering whether a new risk marker should be added to a risk predictor, especially if measuring the marker will impose a real cost (e.g., adding question(s) to a risk assessment tool administered by professional staff). Firstly, does the model already perform sufficiently well that it is difficult to improve its performance? Secondly, is the new risk marker likely to be prevalent rather than rare? Thirdly, does the risk marker provide new prognostic information (a new risk factor) and/or improve the measurement of an existing risk factor or domain, rather than being strongly correlated with risk factors already well-assessed? These should be key considerations when evaluating the option of adding a risk marker to an existing assessment instrument, and – if the answer to any of these questions, or other practical questions such as the cost and reliability of measuring the risk factor, is in doubt – piloting the changes would be highly desirable.

Combining actuarial assessment and clinical judgement.

The existing evidence from OASys is that its clinical judgements of risk of serious harm have considerably weaker predictive validity than OVP scores (Moore, Howard, & Smith-Yau, in press). As the OASys risk of serious harm section is only loosely structured, this accords with meta-analytic results – albeit on samples of sexual offenders - showing that unstructured clinical judgement has weaker predictive validity than other forms of risk prediction (Hanson & Morton-Bourgon, 2009). Nevertheless, at present the design of OASys relegates actuarial risk scores to an advisory role, and it is evident that many assessors pay them little heed. Given the emerging concerns about the clinical rating (Moore et al., in press), forthcoming OASys redesign may lead to actuarial scores being presented as a 'starting point' for risk ratings, which could then be adjusted by assessors. (A purely actuarial system would, at the present time, be unacceptable to assessors and policy makers.) The limited evidence, from three samples of sexual offenders, suggests that clinical adjustment reduces predictive validity (Hanson & Morton-Bourgon, 2009). The OASys redesign may therefore require careful user guidance which attempts to ensure that clinical overrides are restricted to situations where actuarial predictions are genuinely suboptimal (for example, when offenders are suspected of having overseas criminal records; where there is strong evidence of reoffending which has not led to criminal sanction; where the assessment may be compromised by weaknesses of OASys, such as inadequate recording of risk-relevant mental disorders). The validity of the actuarial assessment and the final rating should be monitored, as should the reasons for assessors' decisions to increase or reduce ratings from their actuarial base and the magnitudes and patterns of these changes.

Different types of outcome measure.

This thesis has focused on binary outcomes: that is, whether or not the offender is known to have committed a certain type of reoffence. It includes some consideration (e.g., Table 6 of Chapter 2) of whether serious reoffending can be predicted using the same measures used to predict more violent reoffending. However, alternative and arguably richer measures of reoffending can potentially be used, and would enable further validation of risk measures such as OVP through testing whether predictor scores are associated with such outcomes. PNC data could be used to count frequency of proven reoffending (of whatever type) rather than merely whether it occurred or not. The total cost of proven offending can be estimated by combining official cost-of-crime data with PNC reoffending records (e.g., Jolliffe, Farrington, & Howard, 2013). In both cases, PNC data should be carefully scrutinised to determine whether charging practices vary by location (i.e., whether to charge individuals with all the offences they are suspected of committing or merely focus on the most serious), as this could affect some applications such as comparisons of probation trust performance. A further option is to ask individuals to self-report their reoffending. For sexual offenders in particular, this has been found to result in much greater recidivism rates and frequencies than official data (e.g., Abel et al., 1988). While I agree that this provides valuable information, the conduct of further self-report studies is likely to present considerable challenges given recent increases in the intensity of formal ethical scrutiny of research proposals by both academic institutions and governments.

Conclusion

This thesis has presented substantial new evidence which develops and improves the actuarial prediction of violent and sexual risk, which is a necessary activity for large-scale correctional services seeking to protect the public through the allocation of their limited treatment and risk management resources. It includes a new risk scale for nonsexual violence risk, which has been implemented in prisons and probation services across England and Wales. This risk scale is shown to substantially improve prediction compared with the scales previously available, and to successfully incorporate dynamic risk factors in a way which allows the scale to be productively reviewed over the course of an offender's period of community supervision. Additionally, this scale is validated for the assessment of criminals with a history of sexual offending. Other findings, on specialisation in and within sexual offending, provide important preparatory material for a revised sexual recidivism predictor (which lies outside the scope of this thesis).

This concluding chapter has summarised the above studies, and acknowledged their limitations. It also points the way forward for further research into the prediction of violent and sexual recidivism. While there must be limits to the ability of actuarial risk assessment instruments to identify those most at risk of committing dangerous offences in the future, it is by no means clear that these limits have been reached, and there are still several avenues for new research to explore which could further improve the effectiveness of these important instruments.

REFERENCES

- Abel, G. G., Becker, J. V., Cunningham-Rathner, J., Mittelman, M. S., & Rouleau, J. L. (1988). Multiple paraphilic diagnoses among sex offenders. *Bulletin of the American Academy of Psychiatry and Law*, 16, 153-168.
- Ahlmeyer, S., Heil, P., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 12, 123-138.
- Anderson, C., & Dill, K. (2000). Video games and aggressive thoughts, feelings and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, 78, 772-790.
- Andrews, D.A., & Bonta, J. (1995). *Level of Service Inventory – Revised*. Toronto: Multi-Health Systems.
- Andrews, D.A., & Bonta, J. (2003). *The psychology of criminal conduct*. 3rd edition. Cincinnati, Ohio: Anderson.
- Andrews, D.A., & Bonta, J. (2007). *The psychology of criminal conduct* (4th ed.). Cincinnati, Ohio: Anderson.
- Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, 16, 39-55.

- Andrews, D.A., Bonta, J., & Wormith, J.S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto: Multi-Health Systems.
- Archer, J. (2005). Testosterone and human aggression: a review of the challenge hypothesis. *Neuroscience and Biobehavioural Reviews*, 30, 319-345.
- Archer, J. (2007). Physical aggression as a function of perceived fighting ability among male and female prisoners. *Aggressive Behavior*, 33, 563-573.
- Austin, P.C., & Steyerberg, E.W. (2013). Predictive value of risk factors and markers. A simulation study of the effect of novel markers on different performance measures for logistic regression models. *Statistics in Medicine*, 32, 661-672.
- Babchishin, K.M, Hanson, R.K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, 19, 442-461.
- Babchishin, K.M., Hanson, R.K., & Herman, C. (2011). The characteristics of online sex offenders. A meta-analysis. *Sexual Abuse: a Journal of Research and Treatment*, 23, 92-123.
- Bandura, A. (1977). *Social learning theory*. New Jersey: Prentice Hall.
- Banks, S., et al. (2004). A multiple-models approach to violence risk assessment among people with mental disorder. *Criminal Justice and Behavior*, 31, 324-340.
- Barbaree, H.E., Langton, C.M., Blanchard, R., & Boer, D. (2008). Aging versus stable

enduring traits as explanatory concepts in sexual offender recidivism. Partitioning actuarial prediction into conceptually meaningful components. *Criminal Justice and Behavior*, 36, 443-465.

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

Barratt, E. (1994). Impulsiveness and aggression. In Monahan, J., and Steadman, H.J. (eds.), *Violence and mental disorder: Developments in risk assessment* (pp.61-79). Chicago: University of Chicago Press.

Beech, A. (2008). *Risk assessment in the 21st century. Towards an integrative model of risk in sexual offending*. Presentation to the Risk Management Authority. Available at <http://www.rmascotland.gov.uk/events-and-learning/events/>

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

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

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2008). Forecasting murder within a population of probationers and parolees. A high-stakes application of statistical learning. *Journal of the Royal Statistical Society Series A*, 172, 1-21.

Berkovitz, L. (1993). *Aggression: Its causes, consequences and control*. New York: McGraw-Hill.

Blackburn, R. (1993). *The psychology of criminal conduct: Theory, research and practice*. Chichester, United Kingdom: John Wiley & Sons.

Blumstein, A., Cohen, J., Das, S., & Moitra, S.D. (1988). Specialization and seriousness during adult criminal careers. *Journal of Quantitative Criminology*, 4, 303-345.

Boer, D. (1997). *SVR-20. Manual for the Sexual Violence Risk-20. Professional guidelines for assessing risk of sexual violence*. British Columbia Institute Against Family Violence.

Boles, S.M., & Miotto, K. (2003). Substance abuse and violence: A review of the literature. *Aggression and Violent Behavior*, 8, 155-174.

Bonta, J. (2008). Advances in offender risk assessment. *Public Safety Canada Research Summary*, Vol. 13, No. 2. Available at <http://www.publicsafety.gc.ca/res/cor/sum/cprs200803-eng.aspx>

Brame, R., Bushway, S.D., Paternoster, R. and Thornberry, T.P. (2005). Temporal linkages in violent and non-violent criminal activity. *Journal of Quantitative Criminology*, 21, 149-174.

Britt, C. L. (1996). The measurement of specialization and escalation in the criminal career: An alternative modeling strategy. *Journal of Quantitative Criminology*, 12, 193-222.

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

Burgess, E. W. (1936). Protecting the public by parole and parole prediction. *Journal of Criminal Law and Criminology*, 27, 491-502.

Bushman, B.J., & Anderson, C.A. (2001). Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychological Review*, 108, 273-279.

Bushway, S.D., Nieuwbeerta, P., & Blokland, A. (2011). The predictive value of criminal background checks. Do age and criminal history affect time to redemption. *Criminology* 49, 27-60.

Campbell, M. A., French, S., & Gendreau, P. (2007). *Assessing the utility of risk assessment tools and personality measures in the prediction of violent recidivism among adult offenders*. (Public Safety Canada Research Report, 2007-04.) Available at <http://www.ps-sp.gc.ca/res/cor/rep/risk-tools-200704-eng.aspx>

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

Conference Board of Canada (2009). *The net federal fiscal benefit of CSC programming*. (Research Report 2009 No. R-208). Available at <http://www.csc-scc.gc.ca/text/rsrch/reports/r208/r208-eng.shtml>

Cooke, D. J., & Michie, C. (2010). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior*, 34, 259-274.

Cooke, D.J., & Johnstone, L. (2010). Somewhere over the rainbow. Improving violence risk management in institutional settings. *International Journal of Forensic Mental Health*, 9, 150-158.

Craig, L.A. (2011). The effect on age on sexual and violent recidivism. *International Journal of Offender Therapy and Comparative Criminology*, 55, 75-97.

Davey, L., Day, A., & Howells, K. (2005). Anger, over-control and serious violent offending. *Aggression and Violent Behaviour*, 10, 624-635.

Debidin, M. (Ed.). (2009). *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-09*. (Ministry of Justice Research Series, 16/09.) Available at <http://hlsweb.dmu.ac.uk/pgcpd/roh436/official-documents/Debdin%20Compendium%20of%20OASys%20research.pdf>

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated ROC curves. A nonparametric approach. *Biometrics*, 44, 837-845.

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

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

Dow, E.A., Jones, C., & Mott, J.E. (2005). An empirical approach to recidivism classification. *Criminal Justice and Behavior*, 32, 223-247.

Duwe, G. (2012). Predicting first-time sexual reoffending among prisoners without a prior sex offence history. The Minnesota Sexual Offending Risk Estimate. *Criminal Justice and Behavior*, 39, 1436-1456.

Edens, J. F., & Douglas, K. S. (2006). Assessment of interpersonal aggression and violence: Introduction to the special issue. *Assessment*, 13, 221-226.

Edwards, T.M. (2009). Capacity and the adolescent brain. *Psychiatry, Psychology and the Law*, 16, 427-434.

Eke, A. W., Seto, M. C., & Williams, J. (2011). Examining the criminal history and future offending of child pornography offenders: an extended prospective follow-up study. *Law and Human Behavior*, 35, 466-478.

Elliot, I. A., & Beech, A. R. (2009). Understanding online child pornography use: Applying sexual offense theory to internet offenders. *Aggression and Violent Behavior*, 14, 180-193.

Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (2000, November). *Cross-validation of the Minnesota Sex Offender Screening Tool—Revised*.

Paper presented at the 19th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Farrington, D. (1999). A criminological research agenda for the next millennium.

International Journal of Offender Therapy and Comparative Criminology, 43, 154-167.

Farrington, D.P., Snyder, H.N., & Finnegan, T.A. (1988). Specialization in juvenile court careers. *Criminology*, 26, 461-487.

Gendreau, P., Smith, P., & Thériault, Y. L. (2009). Chaos theory and correctional treatment: Common sense, correctional quackery, and the law of fartcatchers. *Journal of Contemporary Criminal Justice*, 25, 384-396.

Gesch, C.B., Hammond, S.M., Hampson, S.E., Eves, A., & Crowder, M.J. (2002). Influence of supplementary vitamins, minerals and essential fatty acids on the antisocial behaviour of young adult prisoners. *British Journal of Psychiatry*, 181, 22-28.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients to make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53–96.

Gilbert, F., Daffern, M., Talevski, D., & Ogloff, J.P. (2013). The role of aggression-related cognition in the aggressive behavior of offenders. A General Aggression Model perspective.

Criminal Justice and Behavior, 40, 119-138.

Gottfredson, M.R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA., Stanford University Press.

Gottfredson, S.D., & Moriarty, L. J. (2006). Clinical versus actuarial judgments in criminal justice decisions: should one replace the other? *Federal Probation*, 70 (2), 15-18.

Grann, M., & Langstrom, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior*, 34, 22-36.

Haden, S.C., & Scarpa, A. (2007). The noradrenergic system and its involvement in aggressive behaviors. *Aggression and Violent Behavior*, 12, 1-15.

Hahn Fox, B., & Farrington, D. (2012). Creating burglary profiles using latent class analysis. A new approach to offender profiling. *Criminal Justice and Behavior*, 39, 1582-1611.

Hanson, R. (2002). Recidivism and age: Follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence*, 17, 1046-1062.

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

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. (User Report 97-04.) Ottawa, Ontario, Canada: Department of the Solicitor

General of Canada.

Hanson, R.K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, 20, 212-217.

Hanson, R.K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene*, 15 (1), 15-17. Available at:

<http://www.cpa.ca/cpasite/userfiles/Documents/Criminal%20Justice/Crime%20Scene%202008-04.pdf>

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

Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism amongst sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior*, 34, 198-211.

Hanson, R.K., & Howard, P.D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior*, 34, 275-281.

Hanson, R. K., Morton, K. E., & Harris, A. J. R. (2003). Sexual offender recidivism risk: What we know and what we need to know. In R. Prentky, E. Janus, M. Seto, & A. W. Burgess (Eds.), *Annals of the New York Academy of Sciences: Vol. 989. Sexually coercive behavior: Understanding and management* (pp. 154–166). New York: New York Academy of Sciences.

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

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

Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessment for sex offenders*. (User Report 99-02.) Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119_136.

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002*. (Report No. 2003-01.) Ottawa, Canada: Public Safety and Emergency Preparedness Canada.

Harcourt, B.E. (2007). *Against prediction: Profiling, policing and punishing in an actuarial age*. Chicago: University of Chicago Press.

Hare, R.D. (2004). *Hare Psychopathy Checklist – Revised (PCL-R). 2nd Edition*. Toronto: Multi-Health Systems.

Harre, R., & Lamb, R. (1983). *The encyclopaedic dictionary of psychology*. Great Britain: Basil Blackwell Publisher Limited.

Harrell, F.E., Lee, K.L., & Mark, D.B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15, 361-387.

Harris, D. A., Knight, R. A., Smallbone, S., & Dennison, S. (2011). Post-release specialization and versatility in sexual offenders referred for civil commitment. *Sexual Abuse: A Journal of Research and Treatment*, 23, 243-259.

Harris, A.J.R., & Hanson, R.K. (2012, October). *When is a sex offender no longer a sex offender?* Paper presented at the meeting of the Association for the Treatment of Sexual Abusers, Denver, Colorado.

Harris, D. A., Mazerolle, P., & Knight, R. A. (2009). Understanding male sexual offending: A comparison of generalist and specialist theories. *Criminal Justice and Behavior*, 36, 1051-1069.

Harris, D. A., Smallbone, S., Dennison, S., & Knight, R. A. (2009). Specialization and versatility in sexual offenders referred for civil commitment. *Journal of Criminal Justice*, 37, 37-44.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15, 413-425.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (2007, August 15). Abandoning evidence-based risk appraisal in forensic psychiatry: Comments on Hart et al. *British Journal of Psychiatry* (electronic letter). Available at <http://bjp.rcpsych.org/cgi/eletters/190/49/s60#5674>

Harris, G. T., Rice, M. E., & Quinsey, V. L. (2008). Shall evidence-based risk assessment be abandoned? *British Journal of Psychiatry*, 192, 154.

Harris, P.M. (2006). What community supervision officers need to know about actuarial risk assessment and clinical judgement. *Federal Probation*, 70 (2), 8-14.

Hart, H. (1923). Predicting parole success. *Journal of the American Institute of Criminal Law and Criminology*, 41, 405-413.

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

Hart, S.D., & Cooke., D.J. (2013). Another look at the (im-)precision of individual risk estimates made using actuarial risk assessment instruments. *Behavioral Sciences and the Law*, 31, 81-102.

Hart, S.D., Cooke, D.J., & Michie, C. (2007). Precision of actuarial risk assessment instruments: Evaluating the “margins of error” of group v. individual predictions of violence. *British Journal of Psychiatry*, 190, 60-65.

Heil, P., Ahlmeyer, S., & Simons, D. (2003). Crossover sexual offenses. *Sexual Abuse: A Journal of Research and Treatment*, 15, 221-236.

Heilbrun, K., Yasuhara, K., & Shah, S. (2009). Violence risk assessment tools: Overview and clinical analysis. In Otto, R.K., & Douglas, K.S. (Eds.), *Handbook of violence risk assessment* (pp. 1-18). New York: Routledge.

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

Hilton, N.Z., Harris, G.T., & Rice, M.E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist*, 34, 400-409.

HM Prison Service (2009). Generic parole process. (Prison Service Order 6010). Available at http://www.insidetime.org/resources/psd/psd_6010_generic_parole_process.pdf

HM Treasury (2010). *Spending Review 2010*. London: HM Treasury.

Hollin, C. R. (1995). The meaning and implications of 'programme integrity'. In J. McGuire (Ed.), *What works: Reducing reoffending: Guidelines from Research and Practice*. (pp. 195-208). Chichester, England: John Wiley & Sons.

Home Office (2006). *Offender Assessment System OASys User Manual*. (Revised July 2006.) London: Home Office.

House of Commons Justice Committee (2009). *Cutting crime. The case for justice reinvestment*. Available at

<http://www.publications.parliament.uk/pa/cm200910/cmselect/cmjust/94/94i.pdf>

Howard, P. (2006). *The Offender Assessment System: An evaluation of the second pilot* (Home Office Research Findings 278.) Available at

<http://webarchive.nationalarchives.gov.uk/20110218135832/http://rds.homeoffice.gov.uk/rds/rf2006.html>

Howard, P. (2009). *Improving the prediction of reoffending using the Offender Assessment System* (Ministry of Justice Research Summary 2/09). Available at

webarchive.nationalarchives.gov.uk/20110201125714/http://www.justice.gov.uk/publications/docs/oasys-research-summary-02-09.pdf

Howard, P. (2011). Hazards of different types of reoffending. *Ministry of Justice Research Series 3/11*. Available at <https://www.gov.uk/government/publications/hazards-of-different-types-of-reoffending>

Howard, P. (in press). OGP and OVP2. The revised OASys predictors. In Moore, R. (ed.). *A compendium of research on the Offender Assessment System (OASys) 2009-2012*. London: Ministry of Justice.

Howard, P., & Barnett, G. (in press). Development of a new sexual reoffending predictor. In Moore, R. (ed.). *A compendium of research on the Offender Assessment System (OASys) 2009-2012*. London: Ministry of Justice.

Howard, P., Clark, D., & Garnham, N. (2006). *An evaluation and validation of the Offender Assessment System (OASys)*. London: Home Office

Howard, P.D., & Dixon, L. (2011; 2011a in some text). Developing an empirical classification of violent offenses for use in the prediction of recidivism in England and Wales. *Journal of Aggression, Conflict and Peace Research*, 3, 141-154.

Howard, P.D., & Dixon, L. (2012; 2011b in some text). Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys Violence Predictor. *Criminal Justice and Behavior*, 39, 287-307.

Howard, P., Francis, B., Soothill, K., & Humphreys L. (2009). *OGRS 3: The revised Offender Group Reconviction Scale* (Ministry of Justice Research Summary, 7/09). Available at webarchive.nationalarchives.gov.uk/20110201125714/http://www.justice.gov.uk/publications/docs/oasys-research-summary-07-09-ii.pdf

Howard, P., & Moore, R. (2009). *Measuring changes in risk and need over time using OASys* (Ministry of Justice Research Summary, 10/09). Available at webarchive.nationalarchives.gov.uk/20110201125714/http://www.justice.gov.uk/publications/docs/research-summary-oasys-10-09.pdf

Howells, K. (2009). Recent development and future directions in understanding cognition and emotion in offenders: a commentary. *Psychology, Crime & Law*, 15, 285-292.

Hudson, S., & Ward, T. (1997). Relapse prevention: Future directions. *Sexual Abuse: A Journal of Research and Treatment*, 8, 249-256.

Jolliffe, D., Farrington, D. P., & Howard, P. (2013). How long did it last? A 10-year reconviction follow-up study of high intensity training for young offenders. *Journal of Experimental Criminology*, 9, 515-531.

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

Kattan, M. W. (2003). Judging new markers by their ability to improve predictive accuracy. *Journal of the National Cancer Institute*, 95, 634-635.

Kingston, D. A., Yates, P. M., Firestone, P., Babchishin, K., & Bradford, J. M. (2008). Long-term predictive validity of the Risk Matrix 2000: A comparison with the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research and Treatment*, 20, 466-484.

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

Kroner, D.G., Mills, J.F., & Reddon, J.R. (2005). A Coffee Can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law and Psychiatry*, 28, 360-374.

Kroner, D.G., Mills, J.F., Reitzel, L.R., Dow, E., Aufderheide, D.H., & Railey, M.G. (2007).

Directions for violence and sexual risk assessment in correctional psychology. *Criminal Justice and Behavior*, 34, 906-918.

Kropp, R. P., Hart, S., Webster, C., & Eaves D. (1999). *Spousal Assault Risk Assessment Guide User's Manual*. Toronto: Multi-Health Systems, Inc.

Langan, P. A., & Levin, D. J. (2002). *Recidivism of prisoners released in 1994*. Washington, D.C.: US Department of Justice.

Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders. Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior*, 34, 37-59.

Laulik, S., Allam, J., & Sheridan, L. (2007). An investigation into maladaptive personality functioning of internet sex offenders. *Psychology, Crime & Law*, 13, 523-535.

Lehmann, R.J.B., Goodwill, A.M., Gallasch-Nemitz, F., Biedermann, J., Dahle, K.-P. (2012). Applying crime scene analysis to the prediction of sexual recidivism in stranger rapes. *Law and Human Behavior*, 37, 241-254.

Liu, Y.Y., Yang, M., Ramsay, M., Li, X.S., & Coid, J.W. (2011). A comparison of logistic regression, classification and regression tree, and neural networks models in predicting

violent reoffending. *Journal of Quantitative Criminology*, 27, 547-573.

Lussier, P. (2005). The criminal activity of sexual offenders in adulthood: Revisiting the specialization debate. *Sexual Abuse: A Journal of Research and Treatment*, 17, 269-291.

Lussier P., LeBlanc, M., & Proulx, J. (2005). The generality of criminal behavior: A confirmatory factor analysis of the criminal activity of sexual offenders in adulthood. *Journal of Criminal Justice*, 33, 177-189.

Mair, G., Burke, L., & Taylor, S. (2006). 'The worst tax form you've ever seen'? Probation officers' views about OASys. *Probation Journal*, 53, 7-23.

Mann, R.E., Hanson, R.K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, 20, 1-27.

Maruna, S. (2011). Why do they hate us? Making peace between prisoners and psychology. *International Journal of Offender Therapy and Comparative Criminology*, 55, 671-675.

Mazerolle, P., Brame, R., Paternoster, R., Piquero, A., & Dean, C. (2000). Onset age, persistence and criminal versatility. Comparisons across gender. *Criminology*, 38, 1143-1172.

McGloin, J.M., Sullivan, C.J., & Piquero, A.R. (2009). Aggregating to versatility? Transitions among offender types in the short term. *British Journal of Criminology*, 49, 243-264.

McGuire, J. (2004). *Understanding psychology and crime: Perspectives on theory and action*. Berkshire, UK: Open University Press.

Miethe, T. D., Olson, J., & Mitchell, O. (2006). Specialization and persistence in the arrest histories of sex offenders: A comparative analysis of alternative measures and offense types. *Journal of Research in Crime and Delinquency*, 43, 204-229.

Miller, J.D., Lynam, D., & Leukefel, C. (2003). Examining antisocial behavior through the lens of the Five Factor Model of personality. *Aggressive Behavior*, 29, 497-514.

Miller, J.D., Zeichner, A., & Wilson, L.F. (2012). Personality correlates of aggression: evidence from measures of the Five Factor Model, UPPS model of impulsivity, and BIS/BAS. *Journal of Interpersonal Violence*, 27, 2903-2919.

Miller, S.L., & Maner, J.K. (2008). Coping with romantic betrayal: Sex differences in responses to partner infidelity. *Evolutionary Psychology*, 6, 413-426.

Mills, J.F., Kroner, D.G., & Hemmati, T. (2007). The validity of violence risk assessments: An issue of individual item performance. *Psychological Services*, 4, 1-12.

Ministry of Justice (2006). *Implementation of the risk of harm guidance and training resource pack*. (Probation Circular 22/2006.) Available at <http://webarchive.nationalarchives.gov.uk/20101216070244/http://www.probation.homeoffice.gov.uk/output/page31.asp>

Ministry of Justice (2007) [Chapter 1]. *Working in partnership to reduce re-offending and make communities safer: A Consultation*. Available at <http://webarchive.nationalarchives.gov.uk/20100202100532/dcsf.gov.uk/offenderlearning//index.cfm?fuseaction=news.view&newsid=45>

Ministry of Justice (2007) [Chapter 4]. *National Standards for the management of offenders: Standards and implementation guidance*. Available at <http://webarchive.nationalarchives.gov.uk/20100105030830/http://www.probation.homeoffice.gov.uk/files/pdf/OM%20National%20Standards%202007.pdf>

Ministry of Justice (2009a). *Offender caseload management statistics 2008*. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/218068/offender-management-caseload-statistics-2008-2.pdf

Ministry of Justice (2009b). *Reoffending of adults: Results from the 2008 cohort. Appendix A: Statistical tables*. Available at <http://webarchive.nationalarchives.gov.uk/20100406130654/justice.gov.uk/publications/reoffendingofadults.htm>

Ministry of Justice (2010a). *Offender caseload management statistics 2009*. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/218065/omcs-2009-complete-210710a.pdf

Ministry of Justice (2010b). *Multi-Agency Public Protection Arrangements Annual Report*

2009/10. Available at <https://www.gov.uk/government/publications/multi-agency-public-protection-arrangements-annual-report-2009-10>

Ministry of Justice (2011). *Offender management reform and payment by results*. Available at <http://www.justice.gov.uk/offenders/payment-by-results>

Ministry of Justice (2012). *Criminal Justice Statistics: Quarterly update to December 2011*. Available at <http://www.justice.gov.uk/downloads/statistics/criminal-justice-stats/conviction-tables-1211.xls#Table Q4.1!A1>

Ministry of Justice (2013). *Transforming rehabilitation*. Available at <http://www.justice.gov.uk/about/justice/transforming-justice/transforming-rehabilitation>

Modecki, K.L. (2008). Addressing gaps in the maturity of judgment literature. Age differences and delinquency. *Law and Human Behavior*, 32, 78-91.

Moffitt, T.E. (2006). *Taking stock*. New Brunswick, NJ., Transaction.

Monahan, J., Steadman, H. J., Robbins, P. C., Appelbaum, P. S., Banks, S., Grisso, T. et al. (2005). An actuarial model of violence risk assessment for persons with mental disorders. *Psychiatric Services*, 56, 810-815.

Moore, R. (2009). *The internal reliability and construct validity of the Offender Assessment System (OASys)* (Ministry of Justice Research Summary 6/09). Available at webarchive.nationalarchives.gov.uk/20110201125714/http://www.justice.gov.uk/publications/

[docs/oasys-research-summary-06-09.pdf](#)

Moore, R., Howard, P., & Smith-Yau, W. (in press). Reliability and validity of the risk of serious harm ratings. In Moore, R. (ed.). *A compendium of research on the Offender Assessment System (OASys) 2009-2012*. London: Ministry of Justice.

Morton, S. (2009). *Can OASys deliver consistent assessments of offenders? Results from the inter-rater reliability study* (Ministry of Justice Research Summary 1/09). Available at webarchive.nationalarchives.gov.uk/20110201125714/http://www.justice.gov.uk/docs/oasys-research-summary-01-09.pdf

Mossman, D., & Sellke, T. M. (July 5, 2007). Avoiding errors about “margins of error”. *British Journal of Psychiatry* (electronic letter). Retrieved, July 30, 2012 from http://bjp.rcpsych.org/content/190/49/s60/reply#bjprcpsch_el_4856?sid=f06b6af3-df98-4271-a3ba-da91b11ae53f

National Offender Management Service (2010). *Suitability for accredited interventions: Risk, need and responsivity targeting criteria for offender managers and treatment managers and HMPS therapy managers*. Unpublished internal government document.

National Probation Service (2007). *Assessment and management of sex offenders*. (Probation Circular 17/2007.) Available at <http://webarchive.nationalarchives.gov.uk/20101216070244/http://www.probation.homeoffice.gov.uk/output/page31.asp>

National Probation Service (2008). *National rules for tiering cases and associated guidance*.

(Probation Circular 08/08.) Available at

<http://webarchive.nationalarchives.gov.uk/20101216070244/http://www.probation.homeoffice.gov.uk/output/page31.asp>

National Probation Service (2009). *Determining pre-sentence report type*. (Probation Circular

6/09.) Available at

<http://webarchive.nationalarchives.gov.uk/20101216070244/http://www.probation.homeoffice.gov.uk/output/page31.asp>

Ogilvie, J.M., Stewart, A.L., Chan, R.C.K., Shum, D.H.K. (2011). Neuropsychological measures of executive functioning and antisocial behaviour. *Criminology*, 49, 1063-1107.

Olson, C. (2004). Media violence research and youth violence data: why do they conflict?

Academic Psychiatry, 28, 144-150.

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

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

Osgood, D.W., & Schreck, C.J. (2007). A new method for studying the extent, stability and

predictors of individual specialization in violence. *Criminology*, 45, 273-312.

Palmer, E.J, McGuire, J., Hatcher, R.M., Hounsone, J.C., Bilby, C.A.L., & Hollin, C.R. (2009). Allocation to offending behavior programs in the English and Welsh probation service. *Criminal Justice and Behavior*, 36, 909-922.

Parrott, D.J., & Giancola, P.R. (2007). Addressing “The criterion problem” in the assessment of aggressive behaviour. Development of a new taxonomic system. *Aggression and Violent Behaviour*, 12, 280-299.

Petersilia, J. (2003). *When prisoners come home. Parole and prisoner reentry*. New York: Oxford University Press.

Piquero, A. R., Mazerolle, P, Brame, R. & Dean, C. W. (1991). Onset age and offense specialization. *Journal of Research in Crime and Delinquency*, 36, 275-299.

Pitts, S. (2010). Personal communication, 28 October 2010. (At this time, Steve Pitts was NOMS Head of International Relations and External Programmes.)

Prior, D., Farrow, K., Hughes, N., Kelly, G., Manders, G., White, S., & Wilkinson, B. (2011). *Maturity, young adults and criminal justice. A literature review*. University of Birmingham Institute of Applied Social Studies. Available at <http://www.t2a.org.uk/wp-content/uploads/2011/09/Birmingham-University-Maturity-final-literature-review-report.pdf>

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C.A. (1998). *Violent offenders:*

Appraising and managing risk. Washington, DC: American Psychological Association.

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

Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, *10*, 85-105.

Raykar, V. C., Steck, H., Krishnapuram, B., Dehing-Oberije, C., & Lambin, P. (2008). On ranking in survival analysis: Bounds on the Concordance Index. *Advances in Neural Information Processing Systems*, *20*, 1209-1216.

Raynor, P., Kynch, J., Roberts, C., & Merrington, S. (2000). Risk and need assessment in probation services: an evaluation. *Home Office Research Study*, *211*. London: Home Office.

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

Rolison, J.J., Hanoach, Y., & Gummerum, M. (2013). Characteristics of offenders: the HEXACO model of personality as a framework for studying offenders' personality. *Journal of Forensic Psychiatry & Psychology*, *24*, 71-82.

Sample, L. L., & Bray, T. M. (2003). Are sex offenders dangerous? *Criminology and Public Policy*, *3*, 59-82.

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

Seto, M. C., & Eke, A. W. (2005). The criminal histories and later offending of child pornography offenders. *Sexual Abuse: A Journal of Research and Treatment*, 17, 201-210.

Seto, M. C., Hanson, R. K., & Babchishin, K. M. (2011). Contact sexual offending by men with online sexual offenses. *Sexual Abuse: A Journal of Research and Treatment*, 23, 124-145.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis. Modelling change and event occurrence*. Oxford: Oxford University Press.

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

Smallbone, S., & Wortley, R. (2004). Onset, persistence, and versatility of offending among adult males convicted of sexual offenses against children. *Sexual Abuse: A Journal of Research and Treatment*, 16, 285-298.

Smithson, H., Ralphs, R., & Williams, P. (2013). Used and abused. The problematic use of gang terminology in the United Kingdom and its implications for ethnic minority youth.

British Journal of Criminology, 53, 113-128.

Soothill, K., Fitzpatrick, C. & Francis, B. (2009). *Understanding Criminal Careers*.

Cullompton, United Kingdom: Willan.

Soothill, K., Francis, B., Ackerley, E., & Humphreys, L. (2007). Changing patterns of offending behaviour among young adults. *British Journal of Criminology*, 48, 75-95.

Soothill, K., Francis, B., Ackerley, E., & Sanderson, B. (2000). Sex offenders: Specialists, generalists or both? *British Journal of Criminology*, 40, 56-67.

Soothill, K., Francis, B., & Fligelstone, R. (2002). *Patterns of offending behaviour: A new approach*. (Home Office Online Report, 05/02.) Available at <http://webarchive.nationalarchives.gov.uk/20110218135832/http://rds.homeoffice.gov.uk/rds/onlinepubs1.html>

Stander, J., Farrington, D. P., Hill, G., & Altham, P. M. E. (1989). Markov chain analysis and specialization in criminal careers. *British Journal of Criminology*, 29, 317-335.

Steyerberg, E.W. (2009). *Clinical prediction models: A practical approach to development, validation and updating*. New York: Springer.

Sullivan, C. J., McGloin, J., Pratt, T. C., & Piquero, A. R. (2006). Rethinking the “norm” of offender generality. Investigating specialisation in the short term. *Criminology*, 44, 199-233.

Taxman, F. S., & Thanner, M. (2006). Risk, need and responsivity (RNR): It all depends.

Crime and Delinquency, 52, 28-51.

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment.

Sexual Abuse: a Journal of Research and Treatment, 14, 139-154.

Thornton, D. (2007). *Scoring Guide for Risk Matrix 2000.9/SVC*. Available at

<http://www.birmingham.ac.uk/Documents/college-les/psych/RM2000scoringinstructions.pdf>

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M.

(2003). Distinguishing and combining risks for sexual and violent recidivism. *Annals of New York Academy of Sciences*, 989, 225–235.

van der Put, C. E., Deković, E., Stams, G. J. J. M., Van der Laan, P. H., Hoeve, M., & Van Amelsfort, L. (2011). Changes in risk factors during adolescence. Implications for risk assessment. *Criminal Justice and Behavior*, 38, 248-262.

van Kalmthout, A.M, & Durnescu, I. (Eds.) (2008). *Probation in Europe*. Oisterwijk, The Netherlands: Wolf Legal Publishers.

Walters, G. D. (2007). Methodological and statistical advances in correctional and forensic psychology. *Criminal Justice and Behavior*, 34, 1533-1535.

Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of internet child pornography offenders: A comparison with child molesters. *Sexual Abuse: A Journal of Research and Treatment*, 19, 449-465.

Webster, C. D., Douglas, K. S., Eaves, D. & Hart, S. D. (1997). *HCR-20. Assessing Risk for Violence. Version 2*. Vancouver, British Columbia, Canada: Simon Fraser University.

Weinrott, M., & Saylor, M. (1991). Self-report of crimes committed by sex offenders. *Journal of Interpersonal Violence*, 6, 286-300.

Wikstrom, P.-O., & Svensson, R. (2010). When does self-control matter? The interaction between morality and self-control in crime causation. *European Journal of Criminology*, 7, 395-410.

Williams, H.W. (2012). *Repairing shattered lives. Brain injury and its implications for criminal justice*. Report to the Barrow Cadbury Trust. Available at http://www.bctrust.org.uk/wp-content/uploads/2012/10/Repairing-Shattered-Lives_Report.pdf

Wilson, E.O. (1978). *On human nature*. Cambridge, Mass.: Harvard University Press.

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

Yang, M., Liu, Y., & Coid, J. (2010). *Applying neural networks and other statistical models to the classification of serious offenders and the prediction of recidivism* (Ministry of Justice Research Series 6/10). Available at <https://www.justice.gov.uk/downloads/publications/research-and-analysis/moj->

[research/neural-networks-research.pdf](#)

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

Zimring, F. E. (2004). *An American travesty: Legal responses to adolescent sexual offending*. Chicago: University of Chicago Press.

