National Offender
Management Service

A compendium of research and analysis on the Offender Assessment System (OASys)
2009–2013

Robin Moore (editor)
National Offender Management Service

Ministry of Justice Analytical Series
July 2015

The NOMS Research and Evaluation Team supports effective policy development and operational delivery within the National Offender Management Service and Ministry of Justice by providing high-quality social research and statistical analysis. We aim to publish information to add to the evidence-base and assist with informed debate.

Disclaimer

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice (nor do they represent Government policy).

First published 2015

Any enquiries regarding this publication should be sent to us at National.Research@noms.gsi.gov.uk

This publication is available for download at https://www.gov.uk/government/publications

ISBN 978-1-84099-690-6

Acknowledgements

We would like to express our gratitude to colleagues within the OASys Business Team for their guidance regarding the OASys IT system and assessment policies and processes, and to analytical colleagues within the Ministry of Justice (MoJ) and National Offender Management Service (NOMS) for their provision of Police National Computer (PNC) reoffending data, recall data and treatment data. We are further indebted to those assessors and managers who participated in the user perspective study (Chapter 2).

Finally, thanks are due to those who provided constructive feedback on various draft reports as part of the peer review process.

Contents

List of tables

List of figures

Foreword

The Offender Assessment System (OASys) was introduced in 2001, building upon the existing 'What Works' evidence base. It combines the best of actuarial methods of prediction with structured professional judgement to provide standardised assessments of offenders' risks and needs, helping to link these risks and needs to individualised sentence plans and risk management plans.

OASys has improved and helped to join up assessment practice across custody and the community, providing a basis for defensible decision making and supporting the effective management of offenders. By identifying offending-related needs and assisting with the targeting of offenders to interventions, OASys has contributed to reductions in reoffending. OASys data has been used at the local, regional and national levels for resource planning and for segmenting the NOMS caseload, with nearly seven million OASys assessments now collated within a national OASys database.

OASys was piloted prior to implementation and the intention was that, as the evidence base developed, the system would be improved over time. I am therefore pleased to publish this second research compendium produced by the NOMS Research and Evaluation Team which will lead to some further important revisions to OASys, enabling it to continue to play a key role in the delivery and evaluation of interventions and Offender Management. As community rehabilitation services are opened up to a diverse range of new providers, the research findings will inform the continuing development of assessment policies and practices. OASys has been designated an approved tool for use by Community Rehabilitation Companies (CRCs) and the findings have also contributed to the development of a Risk of Serious Recidivism (RSR) tool. The latter is an integral part of the new Case Allocation System (allocating cases to the National Probation System or CRCs), helping to ensure that protection of the public remains paramount and that resources are used as efficiently and effectively as possible.

Michael Spurr
Chief Executive Officer
National Offender Management Service

1. The Offender Assessment System (OASys) and the 2009–2013 research projects

1.1 The development of risk need assessment tools

The effectiveness and efficiency of interventions designed to reduce reoffending and protect the public is dependent upon accurate estimation of offenders' risks and needs. Within England and Wales, as elsewhere, good assessment is thus recognised to be the starting point for managing offenders, as reflected in the acronym of ASPIRE which provides an overview of the management process (Home Office 2005:7):

Assess > **S**entence **P**lan > **I**mplement > **R**eview > **E**valuate

The international reviews of offender assessment tools have identified four key developmental phases (Bonta and Andrews, 2010).

- First generation tools: these rely solely upon subjective, professional judgement.
- Second generation tools: these use actuarial models to predict reoffending, based upon 'static' risk factors such as age and criminal history. A wide range of studies have found that such actuarial instruments outperform clinical or professional judgments when making predictions.
- Third generation risk-need tools: these measure dynamic as well as static risk factors, enabling interventions to be directed to these dynamic factors and changes in offender profile to be monitored.
- Fourth generation tools: these integrate other offender-specific factors important to treatment and enable intervention delivery to be planned and monitored.

Within England and Wales, the fourth generation tool used with adult offenders is the Offender Assessment System (OASys). A separate tool is used for young offenders aged 10–17 known as *Asset* (Baker, 2004; Wilson and Hinks, 2011), while other tools are used for specific types of offending, e.g. Risk Matrix 2000 (RM2000) for sexual offending (Thornton, 2007) and the Spousal Assault Risk Assessment Guide (SARA) for domestic violence (Kropp *et al.*, 1995). For adult offenders who are not assessed through OASys, a second generation tool is available – the Offender Group Reconviction Scale v.3 (OGRS3). This tool predicts proven reoffending with one and two years using age at sentence, gender, number of previous sanctions, age at first sanction and current offence (Howard *et al*., 2009).

Chitty (2004:75) has highlighted the need for any assessment system to include "sufficiently reliable, valid and sensitive measures of risk factors so that they can perform their assessment and monitoring tasks effectively". A more detailed list of the desirable characteristics of a risk and needs assessment system have been set out by Bonta *et al.* (2001:233) – see Table 1.1 below:

Table 1.1 The desired characteristics of risk-needs classification (Bonta et al. 2001, p 233)

Characteristic	Description
Objective	Items described with publicly observable referents; structured administration and scoring rules.
Internal reliability	Items relate to each other and the total score.
Inter-rater reliability	High agreement among test administrators; items are scored the same way producing similar results.
Meaningful	Information makes sense; items consistent with the research on the prediction of recidivism.
Predictive validity	Scores predict relevant outcomes (e.g. recidivism, prison misconduct, parole violation).
Dynamic validity	Changes in scores predict changes in outcome.
Socially unbiased	Items do not violate constitutional / charter rights (e.g. ethnicity, gender).
Generalisation	Instrument applies well to other groups and settings beyond the initial construction sample.

Merrington (2004) has emphasised that ‘a balance has to be struck between technical performance and fitness for purpose’. Further research (e.g. Aubrey & Hough, 1997; Aye Maung & Hammond, 2000) has focused upon the desirable characteristics of an assessment tool from a practitioner perspective, identifying the following key points:

- ***Face validity:*** It must be clear why each item is included.
- ***Clear definitions:*** Clear and unambiguous definitions of the items are required for consistency.
- ***Simple scoring system:*** Question scales that stretch beyond five points have been found to be difficult.
- ***Evidence boxes:*** Practitioners should have the opportunity to express their concerns and elaborate on their assessment.
- ***Offender input:*** Offenders should be provided with the opportunity to express their views.
- ***Useable within limited time constraints:*** Resource implications need to be considered.
- ***Complements current practice:*** The instrument needs to be continually developed.

1.2 The development of OASys

OASys was developed through three pilot studies running from 1999 to 2001 (Howard, Clark and Garnham, 2006). An electronic version of the tool was then rolled-out across both the prison and probation services, with a new single system being implemented in 2013 through the OASys-R project. The value of the tool has been summarised as follows: “OASys is a central part of evidence-based practice. It is designed to be an integral part of the work which practitioners do in assessing offenders; identifying the risks they pose, deciding how to minimise those risks and how to tackle offending behaviour effectively. OASys is designed to help practitioners make sound and defensible decisions” (Home Office, 2002). More specifically, OASys is designed to:

1) assess how likely an offender is to reoffend;
2) identify and classify offending-related needs;
3) assess risk of serious harm, risks to the individual and other risks;
4) assist with managing the risk of serious harm;
5) link the assessment to the sentence plan;
6) indicate the need for further specialist assessments; and
7) measure change during the offender's sentence.

To fulfil these functions, OASys has several different components. The core assessment classifies offending-related needs, encompassing individual-level factors, in terms of 'internal' disposition, personality, reasoning and temperament, and 'external' social or societal factors and their influences on offending behaviour. Selected questions from these sections, alongside offending history and offender demographic information, contribute to two predictors of reoffending: the OASys General reoffending Predictor v.1 (OGP1) and the OASys Violence Predictor v.1 (OVP1). The OGRS3 predictor (based on static factors only) is also calculated. A separate Risk of Serious Harm (RoSH) component focuses upon the likelihood of life-threatening and/or traumatic events, requiring assessors to make informed judgements regarding the risks to various groups (children/public/known adult/staff). Practitioners are thus able to prioritise public protection issues, identifying appropriate requirements, conditions and controls for managing specific risks. The OASys summary sheet utilises information from the core assessment to score the predictors of reoffending and present summaries of offending-related needs, and present summary RoSH information. A sentence plan and risk management plan is developed to address these risk and needs.

A continuing research programme has assisted in the development of OASys over time, helping to ensure that it remains a valid and reliable system. Notably, significant improvements were made to the measurement of offenders' risks and needs in August 2009. The research underpinning these improvements was published within the first OASys research compendium (Debidin, 2009). The chapters within the first compendium covered construct validity, internal reliability, inter-rater reliability, predictive validity and dynamic validity, as well as including analysis of OASys rates of completion, coverage and representativeness, textual analysis, and a review of the underlying evidence.

The analysis of coverage and representativeness was required because while OASys is now in general use, it is not required to be used with all offenders. At the Pre-Sentence Report (PSR) stage, all 15-day adjourned reports must be based on an OASys assessment, but on-the-day, five-day and oral reports can be based upon an OGRS score and an OASys RoSH screening (National Offender Management Service, 2011a). Post-sentence, an assessment should be completed in the community for all those cases designated at Offender Management Tier 2 and above, with the exception of those

Tier 2 cases in which there is a stand-alone unpaid work requirement.[1] In the prison establishments, all 18–20 year olds and all older offenders serving a custodial sentence of at least 12 months should be assessed (National Offender Management Service, 2008).[2]

The analysis presented in the last compendium found that the use of OASys is consistent with the expectation that resources should follow risk – those offenders with an OASys were more likely to have committed a violent offence and to have a high likelihood of reconviction than non-assessed offenders (Moore, 2009a). The ability to validate OASys for use with lower risk offenders is thus restricted.

Supporting the evidence base and 'What Works' principles

A reliable and valid assessment tool is needed to support the existing evidence base on 'What Works' in reducing reoffending, particularly the risk, need and responsivity (RNR) principles (McGuire, 1995). The risk principle ensures that (rehabilitative) interventions are offered to moderate and high risk cases with low risk cases receiving minimal intervention, while the need principle ensures that criminogenic needs are the focus of targeted interventions, rather than other needs which are not related to offending behaviour. The responsivity principle encompasses both general and specific responsivity. While general responsivity promotes the use of cognitive social learning methods to influence behaviour, specific responsivity provides that interventions should be tailored to, amongst other things, the strengths of the offender.

The importance of the RNR principles are highlighted by Bonta and Andrews (2007). They found that recidivism increased if there was a failure to adhere to any of the RNR principles, i.e. if treatment targeted non-criminogenic needs of low risk offenders using non-cognitive-behavioural techniques. In contrast, adherence to all three RNR principles led to a 17% positive difference in average recidivism between treated and non-treated offenders when delivered in residential/custodial settings, and a 35% difference when delivered in community settings.

Supporting the risk principle

To support the risk principle, identifying which offenders should receive the available interventions, OASys includes two robust predictors of reoffending: OGP1 and OVP1. The introduction of OVP1 greatly improved prediction of violence against the person, weapons, robbery, criminal damage and public order ('violent-type') offences and OGP1 improved prediction of other non-sexual ('general')

---

1 Probation Circular 08/2008 sets out the four Offender Management Tiers and how they are to be applied (National Offender Management Service, 2008). The four tiers represent different levels of intervention, creating four broad categories of case: Tier 1 = Punish; Tier 2 = Punish and Help; Tier 3 = Punish and Help and Change; Tier 4 = Punish and Help and Change and Control. Offenders are allocated to the four tiers according to their risks, needs and circumstances. Consequently, the lowest risk offenders receive basic levels of intervention at Tier 1 and the highest risk offenders receive the most intensive interventions at Tier 4.

2 Following the introduction of 'layered OASys' in August 2009, some OASys-eligible offenders receive a full assessment, whilst some receive a condensed standard assessment.

offences. Both predictors use static and dynamic risk factors, based upon analyses of which OASys questions were most strongly associated with reoffending (Howard, 2009).

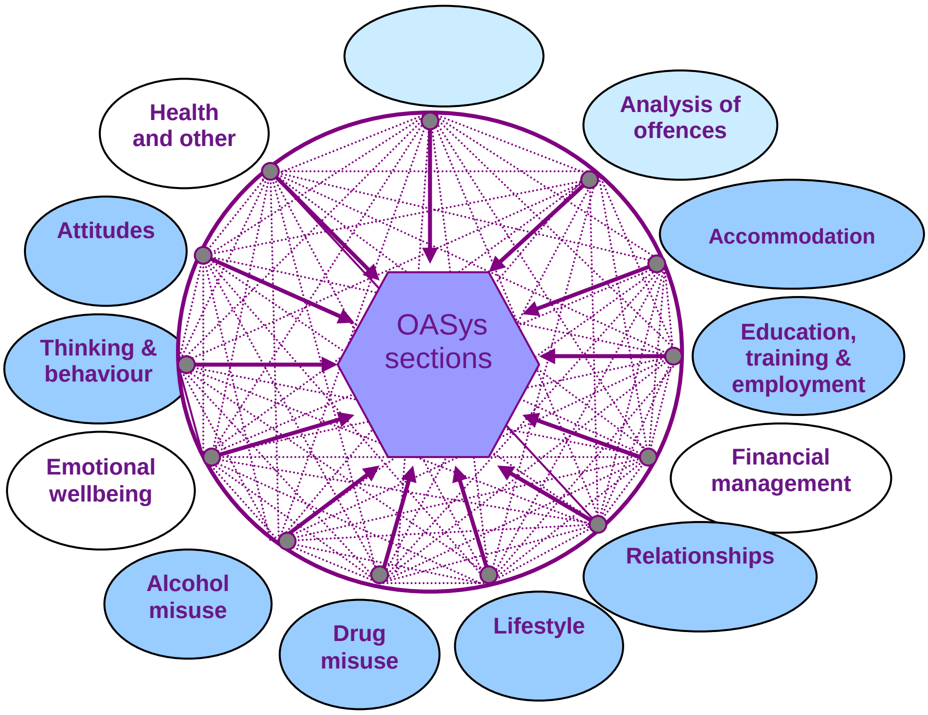
Supporting the need principle

Reviews of the literature have identified the following seven key criminogenic needs (Andrews and Bonta, 2010):

- Procriminal attitudes (thoughts, values and sentiments supportive of criminal behaviour).
- Antisocial personality (low self-control, hostility, adventurous pleasure seeking, disregard for others, callousness).
- Procriminal associates.
- Social achievement (education, employment).
- Family/marital (marital instability, poor parenting skills, criminality).
- Substance abuse.
- Leisure/recreation (lack of prosocial pursuits).

Alongside criminal history, the first three have been termed the “big four” risk factors for recidivism, with the remaining four “moderate” risk factors completing the “central eight” risk factors. Within the main body of OASys, there are ten separate sections, eight of which are now scored as criminogenic needs – see Figure 1.1. The questions contributing to the criminogenic need scores have been streamlined, resulting in clearer distinctions between the differing risk factors, while the cut-off points have been closely aligned to reoffending rates (Moore, 2009b) – see Appendix A for a list of all the scored questions (including which are scored in OGP 1 and OVP1).

Figure 1.1: Construction of the core OASys assessment



Key: Offence/offending information
Scored criminogenic needs
Non-scored sections

Supporting the responsivity principle

To support specific responsivity, assessors are asked to record positive factors in the evidence boxes for each section of the core OASys assessment. The sentence plan within OASys then enables assessors to record 'positive factors to be maintained or developed'. Personal strengths can thus be integrated into the delivery of interventions.

Engaging offenders

Desistance research (e.g. McNeill and Weaver, 2010) has emphasised the importance of engaging with offenders, recognising their individuality and focusing upon the development of positive relationships. To engage offenders within the assessment process, the OASys Self Assessment Questionnaire (SAQ) provides offenders with the opportunity to comment upon their lives (Merrington, 2004). These views can then be considered as part of the interview with the offender and be incorporated within the sentence plan. Analysis has revealed that all the SAQ questions have associations with reoffending (Moore, 2009c). Offenders are also able to comment upon their sentence plan and consider whether they agree with its content. It is recognised good practice that a copy of the final assessment should be shared with the offender.

Using collated assessment data

Completed assessments provide large amounts of standardised information about offenders while they are being supervised. The potential benefits are wide-ranging:

> *Once it is automated and in general use, OASys has the capacity to provide valuable management information, some of which will be used by practitioners to develop profiles of the offenders they are working with and to evaluate overall outcomes. Information will also be of use to local managers, to enable them to identify which risk factors are most common within their local offender population and to help ensure that adequate provision has been made for them. When applied on a national basis, OASys will provide a profile of offenders and their needs, and will permit resources to be allocated effectively*
>
> (Home Office, 2002:3-4)

Now that OASys is both automated and in general use, the collated data has been used widely. For example, analysis of 2008 assessments (O-DEAT, 2009) found the following differences between offender sub-groups:

- Female offenders had relatively high levels of need for relationships and emotional wellbeing, and relatively low levels of need for both thinking/behaviour and attitudes.
- The youngest offenders (aged 18–20) had relatively high levels of need for education, training and employability (ETE) and lifestyle/associates, and a relatively low level of need for emotional wellbeing. The oldest offenders (aged over 40) had relatively low levels of need for ETE and drug misuse.
- Early-onset offenders and the most persistent offenders had relatively high levels of need across the majority of the OASys sections.

More recent OASys figures are presented in a summary of evidence on reducing reoffending (Ministry of Justice, 2014), whilst the OASys data has also been used to segment the NOMS caseload to supply commissioners and providers with standardised offender profiles for differing sub-groups (National Offender Management Service, 2013). The OASys information is a key data source for large-scale research projects, for example the Offender Management Community Cohort Study (Wood *et al*., 2013). OASys data is also used in outcome studies evaluating the effectiveness of accredited interventions (e.g. Sadlier, 2010) and the use of differing sentence requirements (e.g. Bewley, 2012). By the end of March 2014, almost seven million prison and probation assessments had been collated within the central O-DEAT (OASys Data, Evaluation and Analysis Team) database for over one million offenders.

1.3 Research in this compendium

The first OASys research compendium (Debidin, 2009) presented the findings from research and analysis completed between 2006 and 2009. This follow-up compendium covers the studies completed between 2009 and 2013, including a systematic review of the underlying evidence-base, a survey of assessors' views and experiences, and analyses of various aspects of construct validity, internal reliability, predictive validity and dynamic validity. For those studies that have utilised OASys and Police National Computer (PNC) data, a couple of limitations are worthy of note. Firstly, due to the targeting of OASys, the ability to validate the tool for all types of low risk offender is restricted. Secondly, as the PNC data reflects proven reoffending which has led to a formal caution or conviction, it will under-record actual offending behaviour and will be affected by the activities of practitioners within the criminal justice system.

A summary of each chapter within this compendium and the key points are set out below. The focus of Chapters 2 to 7 is on the performance of the current version of OASys, while Chapters 8 to 12 focus upon potential revisions to OASys. A glossary of commonly used terms and concepts can be found towards the end of the compendium – this glossary was included to assist readers and to avoid unnecessary repetition across the chapters.

Section 1: Assessment of current tool

Chapter 2 presents the findings of a study which aimed to capture prison and probation assessors' views and experiences of OASys, highlighting potential improvements. The views of over 1,000 prison and probation assessors were obtained through an online self-completion questionnaire. Twelve follow-up interviews were conducted with OASys leads/managers, exploring issues raised through the online survey. Key findings, and consequent implications, were as follows:

- Approximately nine out of ten assessors felt that the information recorded in an OASys assessment supported them well in managing offenders' risks and needs. Approximately two thirds thought that the level of detail and content at each of the OASys layers and within fast reviews was about right. There was, however, some support for removing areas of duplication and for revisiting the structure and content of the self assessment questionnaire and sentence plan.
- Feedback on the RoSH ratings was positive, providing support to their use in the new Case Allocation System (CAS; National Offender Management System, 2014) for allocating cases to the National Probation System (NPS) or Community Rehabilitation Companies (CRCs).
- More than nine out of ten of the assessors agreed that they regularly used their professional judgement to complete an OASys assessment. Seven out of ten felt that the amounts of professional judgement required to complete an assessment were about right, but approximately one quarter stated it was too little.
- There were some clear resource issues, with over half of the assessors disagreeing that they usually had sufficient working hours to complete assessments. Future proposals for changes to the targeting and content of OASys will need to pay careful attention to the potential resource implications.
- More than four fifths of the assessors felt there was scope for improving the targeting of OASys and its layers. Of those who held this view, around one quarter felt that the full

layer was targeted at too many offenders, while one fifth believed that fast reviews were targeted at too few offenders.

- Just under half of the assessors thought that there was scope for improving the OASys online help. Issues around both navigation and content were raised.
- Approximately two thirds of the assessors felt that the level of quality assurance for OASys assessments was about right, and over half thought that the introduction of the quality assurance process had raised the quality of OASys assessments. However, around three quarters believed that there remained scope for improving quality.
- Areas of concern differed between prison and probation assessors. For example, prison assessors were more likely to say that they did not always have sufficient offender information when completing assessments (50% vs. 29%) and that OASys training was not available when it was needed (53% vs. 36%). Probation assessors were more likely to say that there was too little offender engagement in the sentence planning process (30% vs. 18%) and that unnecessary information was being recorded (31% vs. 14%).
- Areas where awareness needed to be raised included: (i) the value and workings of the actuarial reoffending predictors; (ii) the current targeting criteria for the OASys layers and fast reviews; and (iii) the availability of the online help.

Chapter 3 sets out the findings from research which tested the ability of OGP1 and OVP1 to predict proven reoffending for offenders of different gender, ethnicity and age. Key findings were as follows:

- Among all offenders, actual (proven) reoffending was significantly below the predicted rate, especially for non-violent offending, reflecting known overall reductions in reoffending since OGP1 and OVP1 were created.
- Among women, non-violent reoffending was 3.7% below predicted, compared with 2.1% below predicted for male offenders. While the non-violent reoffending of White offenders was 2.6% below predicted, for Asian and Black offenders it was 3.4% and 2.2% respectively above predicted. Actual and predicted non-violent reoffending were identical for offenders aged 18–19, but actual rates were between 1% and 4% lower than predicted for all other age groups.
- Actual/predicted differences by gender and ethnicity were far smaller for violent reoffending, while violent reoffending rates were 2% above predicted at age 18–19 and at least 4% below predicted for 22–23, 46–50 and 51+ year olds.
- Both predictors achieved reasonable relative predictive validity – successfully distinguishing likely reoffenders and likely non-reoffenders – for all offender groups. Relative predictive validity was greater for female than male offenders, for White offenders than offenders of Asian, Black and Mixed ethnicity, and for older than younger offenders. After controlling for differences in risk profiles, lower validity for all Black, Asian and Minority Ethnic (BME) groups (non-violent reoffending) and Black and Mixed ethnicity offenders (violent reoffending) was the greatest concern.

- Statistical modelling suggests that some between-group differences in risk factors for reoffending may exist, but incorporating these factors does not improve prediction of reoffending sufficiently to justify the introduction of separate predictors. Revision of the predictors will be required to ensure that they reflect contemporary patterns of reoffending.

The study reported in Chapter 4 examined whether scores on supposedly dynamic risk factors changed over the course of probation supervision, and whether changes in risk factor and predictor scores were associated with changes in reoffending risk. Key findings were as follows:

- Mean OGP1 and OVP1 scores fell over the course of offenders' supervision. Scores fell more for non-reoffenders than reoffenders, even though non-reoffenders had lower initial scores. Accommodation, drug misuse and alcohol misuse scores were especially dynamic, with the greatest net reduction being in alcohol misuse, though two OVP1 risk factors did not demonstrate dynamic properties.
- Prediction of reoffending was improved by accounting for changes in dynamic risk, by using current rather than initial assessments. Changes in most OGP1/OVP1 risk factors contributed incrementally to the prediction of reoffending.
- These findings demonstrate the value of reviewing OASys assessments during probation supervision. Reviewing assessments improves prediction of reoffending by keeping dynamic risk factors up to date, and offers an evidence-based mechanism for gradual reductions in the resources allocated to a case. When designing the next iterations of the reoffending predictors, a methodology should be used which accounts for changes in dynamic risk factor scores.

In the study reported in Chapter 5, patterns of reoffending for six types of rare, harmful offence were analysed in order to determine whether offenders specialise in these offences. Key findings were as follows:

- Some degree of specialisation was found for all six offence types. For arson, child neglect, dangerous driving, kidnapping and racially aggravated offending, those with a history of such offending were three to four times more likely to commit further offences than those without, rising to eight times for blackmail. Offence-specific history was therefore a risk factor for all six offence types.
- Arson, kidnapping and racially aggravated offences were well predicted by OVP1. They should be included in the set of offences which OVP classes as violent.
- Child neglect offences were most likely to be committed by young women living with children, especially those with high scores on dynamic risk factors included in OGP1. The principal dynamic risk factor in OGP1 is drug misuse.
- Dangerous driving offences were most likely to be committed by young men with employability, lifestyle and impulsivity problems and histories of driving whilst disqualified or uninsured and/or dangerous driving. OGRS3 and OGP1 were reasonable predictors.

- Blackmail was an extremely rare offence, and the likelihood of such reoffending may be assessed using OGRS3 or OGP1.
- While most of these offences were relatively rare, rates of reoffending among those most at risk were sufficiently high that the possibility of such offending should be explicitly considered when conducting risk assessments in these cases.

Chapter 6 presents the findings of a study which examined the positive, promotive and protective factors recorded within OASys. Positive factors were deemed to be 'promotive' when they were negatively correlated with reoffending, having controlled for risk factors. They were deemed to be 'protective' when moderating the impact of specific risk factors. Key findings were as follows:

- The textual analysis revealed that the positive factors recorded within the OASys sentence plan correspond to the socio-economic and individual-level domains covered by the core OASys assessment. The prevalence rates of the extracted positive factor categories were relatively low, indicating that the full range of positive factors may not always have been considered.
- The optimum model for predicting reoffending included the OGRS3 score, six dynamic risk factors, five promotive factors and one protective interaction. The identification of overlapping risk and promotive processes indicates that, where risk factors are hard to change, interventions can potentially offset the risks of further offending by enhancing promotive factors, assisting with offender engagement. The interaction in the model indicated that positive family relationships moderated the impact of problematic drug misuse.
- The model combining static risk factors, dynamic risk factors, promotive factors and protective interactions performed only marginally better than a model combining static and dynamic risk factors alone. Bearing in mind that OGP1 and OVP1 have high predictive validity, this finding suggests that little would be gained, in terms of accurately predicting reoffending, from a scoring system which distinguished risk factors from promotive/protective factors.
- Consideration should be given to: (i) highlighting further the importance of identifying positive as well as risk factors during OASys assessors' training; (ii) ensuring that the recording of positive factors is carefully monitored through existing quality assurance procedures; (iii) introducing fixed response categories to encourage more systematic recording of positive factors; and (iv) distinguishing between positive factors that need to be maintained and those that need to be developed, assisting in the identification of immediately promotive/protective factors and enabling changes in status (development vs. maintenance) to be monitored.

Chapter 7 presents an analysis of the reliability and validity of OASys Risk of Serious Harm (RoSH) ratings. Probation assessments completed between 2005 and 2008 were analysed and those completed by mid-July 2006 were matched with 24-month reoffending data. Key findings were as follows:

- The analysis revealed not only variation between probation areas in high/very high RoSH prevalence rates, but also differences in the actual minus predicted residual rates (using a checklist to identify offenders likely to be rated as high/very high RoSH). The majority of probation areas had significantly fewer high/very high risk offenders than predicted, with the large urban areas of Merseyside, London, Greater Manchester, West and South Yorkshire all having more high/very risk offenders than predicted.
- There was also considerable variation between probation areas in the use of the RoSH screening overrides. Notably, use of the exemption from full analysis clause ranged from 2% to 26%. Bearing in mind that (i) every exemption has to be clearly evidenced, (ii) the assessor must be confident that the offender is not likely to cause serious harm and (iii) the exemption has to be countersigned by a senior practitioner, the higher rate may be viewed as higher than expected. Greater consistency could be encouraged through improved guidance regarding the use of the overrides and possibly the introduction of structured response options.
- Looking at half-year periods from 2005 to 2008, the increase in high/very high RoSH ratings from 5.8% to 9.9% appeared broadly justified by the characteristics of the offenders who were assessed. The more sizeable shift was from low to medium RoSH ratings; the former falling from 62% to 36%.
- Grave reoffences were predicted with much greatly validity by an actuarial risk assessment score than by the clinical RoSH ratings. It is therefore likely that public protection could be improved by increasing the influence of actuarial scores upon RoSH ratings. As highlighted in Chapter 13, this has led to the development of a new actuarial Risk of Serious Recidivism (RSR) tool (see Appendix H) which is being used alongside the RoSH ratings in the Case Allocation System (CAS) for routing cases to the NPS or to CRCs.

Section 2: Revisions to OASys

Chapters 8 to 10 focus on the development and validation of new static and static/dynamic actuarial predictors of reoffending, covering general, violent and sexual reoffending. Findings from the earlier chapters feed into the development of these predictors. Chapter 8 presents version 4 of the static predictor OGRS, setting out the following key points:

- OGRS4 includes models for general (i.e. all recordable) and violent proven reoffending, known as OGRS4/G and OGRS4/V respectively. In predicting general reoffending, OGRS4/G was found to significantly outperform OGRS3. In predicting violent reoffending, OGRS4/V significantly outperformed OGRS4/G and other operational predictors.

- The new models included an ‘offence-free time’ element, recognising that an offender’s probability of future proven reoffending falls with time after community sentence or discharge from custody without yet reoffending (see Chapter 4 points above). The models thus allow a more accurate comparison of offenders at different stages of community supervision, assisting with the targeting of supervision and treatment resources.
- The improvements in the prediction of both general and violent reoffending were due to the application of offence-free time and other innovations in the coding of risk factors. The choice of ‘primary’ static risk factors – those which must be entered by practitioners – was nevertheless constrained to ensure that all could be coded quickly. The refinements to the coding of the ‘secondary’ risk factors, calculated from the practitioner-entered information, illustrate the degree of fine-tuning required to achieve incremental improvements in the prediction of proven reoffending.
- The nature of the sample used to create OGRS4 means that the new predictors have scope to be used in settings where OGRS3 is not currently used, among offenders with cautions or absolute/conditional discharges from court, and in youth justice. Such use would require the development of user guidance and possibly training. Users who are already familiar with OGRS3 could be issued with more limited guidance covering the improved validity, revisions to offence categories, the offence-free time element, and any subsequent revisions to risk groupings.

Chapter 9 reports on the development of version 2 of OGP and OVP. The chapter sets out the following key points:

- Following feedback from OASys users, the second iteration of OGP predicts all proven reoffending.
- OGP2 and OVP2 have the same static risk factors as those used in OGRS4/G and OGRS4/V, although these factors are scored differently.
- As with OGRS 4, the new models include an ‘offence-free time’ element.
- Dynamic risk factors in both predictors include accommodation, employability, intimate partner relationships, the type of drug used, alcohol misuse, impulsivity, temper control and problem solving skills. OGP2 also includes frequency of drug misuse and pro-criminal activities and attitudes.
- Improvements in the prediction of both general and violent reoffending resulted from the application of offence-free time and other innovations in the selection and coding of risk factors. Version 2 better distinguishes reoffenders from non-reoffenders, and better calibrates actual and predicted reoffending rates for certain offender groups: the highest- and lowest-risk, the oldest and youngest, and females (see Chapter 3 points above).
- The predictors could be introduced in a revision to OASys, accompanied by user guidance.

Chapter 10 examines whether OASys and criminal history information can be combined into a score which improves prediction of the sexual offences most likely to cause serious harm: those involving direct contact with victims. These ‘contact offences’, involving direct and serious harm, include rape, sexual assault, gross indecency, incest, unlawful sexual intercourse and grooming. While clearly still harmful, the sexual offences excluded from this category principally comprise those related to indecent images of children and exhibitionism (e.g. indecent exposure). The chapter sets out the following key points:

- Sexual offences were divided into four groups: contact adult, contact child, paraphilia (e.g. indecent exposure) and indecent images of children. Examining patterns of previous sanctions (i.e. cautions and convictions) and reoffending, for approximately 15,000 offenders, showed that offenders tend to strongly specialise by committing particular types of sexual offence.
- A new seven-item predictor, the OASys Sexual reoffending Predictor (OSP), was developed. This predictor uses static risk factors only and can thus be scored on the basis of summary printouts of individual offenders’ demographics and criminal histories. It was found to be superior to RM2000/s as a predictor of contact sexual reoffending.
- The risk factors in OSP are (strongest first): contact adult sanctions; current age; age at last sexual offence; contact child sanctions; paraphilia sanctions; not first-time entrant; and stranger victim of current sexual offence.
- As OSP has the potential to improve prediction of those sexual offences most likely to cause serious harm, and is no more complex to administer, it is recommended that its implementation in NOMS and police practice should be considered – it has already been incorporated within the new RSR tool and used for segmenting the NOMS caseload. If OSP is fully implemented as a stand-alone predictor, amendments to user guidance and IT systems will be necessary.

While Chapters 8 to 10 focus upon predictors of differing types of reoffending, Chapter 11 shifts attention to the OASys measurements of discrete criminogenic needs, ensuring adherence to the ‘What Works’ criminogenic need principle as well as the risk principle. Key findings were as follows:

- The underlying factor structure of the scored OASys questions corresponds to the eight OASys criminogenic need sections.
- To maximise the item-scale correlations and the dynamism of the scales, as well as alignment to OGP2/OVP2, the analysis supports some changes to the questions which are scored – three questions being removed from the scoring and three being added.
- These amendments leave 31 scored questions across the eight criminogenic need scales, but all the scales (bar one – lifestyle and associates) now have four questions and a 0–8 scale.
- The revised scales were found to be independently associated with reoffending, with some changes required to the criminogenic need cut-off points.

- The above changes have an impact upon the criminogenic need prevalence rates across five of the scales (although a relatively small change for two of the scales) – adjustments in the allocation of resources would be required to ensure that interventions were available to address the revised criminogenic need levels.
- Fewer scales were found to be independently predictive for the BME sub-groups, akin to the lower OGP1 predictive validity for BME offenders reported in Chapter 3.

Chapter 12 presents a systematic review of the literature on the dynamic risk and protective factors for general and violent reoffending, recognising that OASys must not only continue to pass stringent reliability/validity performance criteria but must also continue to reflect the research literature on which it is based. Thirty-two UK and international studies published between January 2000 and November 2011 (heterogeneous in terms of populations, methodology and data reporting) were included in the review, the key findings from which were as follows:

- No new risk domains were identified that would be worthwhile additions to OASys.
- In terms of more specific items within the domains, not all items were consistently identified and those that were most consistently identified matched closely to specific OASys questions.
- Gang membership, which is not currently recorded within OASys, was found to be predictive of future violent reoffending in one relatively large US study. Consideration could thus be given to including a question on gang associations/activities within the current lifestyle and associates section.
- Further reviews of the literature could be undertaken using the same systematic approach, helping to ensure that offender assessment policy within NOMS continues to reflect the most up-to-date knowledge about risk and protective factors.
- There is a clear need for further studies identifying: (i) positive factors which are negatively correlated with reoffending as well as those which moderate the impact of specific risk factors; and (ii) whether there are differences between the dynamic risk and protective factors according to age, gender and ethnicity. Further attention also needs to be given to which dynamic factors are truly causal, where changes over time are associated with changes in future offending behaviour when other factors are held constant.

Chapter 13 focuses upon the key implications from the totality of the research presented in the previous chapters, as well as summarising the work undertaken during 2013 to further validate and recalibrate the actuarial predictors of reoffending – OGRS4 (Chapter 8), OGP2 and OVP2 (Chapter 9), and OSP (Chapter 10). The chapter also sets out how the research recommendations are being taken forward by NOMS, including the design and implementation of the actuarial RSR tool and its use alongside the RoSH ratings for allocating cases to the appropriate community providers (NPS or CRCs). By structuring practitioners' judgements in this way, the intention is to ensure that the most appropriate high risk cases remain with the NPS. Finally, the chapter sets out potential future research, recognising that the validation of a fourth generation assessment tool such as OASys should be seen as on-going so that it reflects developments in the underlying evidence-base, the latest validation methodologies and changes in reoffending patterns, while continuing to support practitioners and current operational priorities and practices. Now that OASys has been designated as an approved tool for use by CRCs, continuing research and validation will enable it to remain fit for purpose across custody and community settings.

2. Prison and probation assessors' views and experiences

This chapter presents the findings of a study capturing prison and probation assessors' views and experiences of OASys. The views of over 1,000 prison and probation assessors were obtained through an online self-completion questionnaire. Twelve follow-up interviews were conducted with OASys leads/managers, exploring issues raised through the online survey. Key points are as follows:

- Approximately nine out of ten assessors felt that the information recorded in an OASys assessment supported them well in managing offenders' risks and needs. Approximately two thirds thought that the level of detail and content at each of the OASys layers and within fast reviews was about right. There was, however, some support for removing areas of duplication and for revisiting the structure and content of the self assessment questionnaire and sentence plan.
- Feedback on the RoSH ratings was positive, providing support to their use in the new Case Allocation System for allocating cases to the National Probation System (NPS) or Community Rehabilitation Companies (CRCs).
- More than nine out of ten of the assessors agreed that they regularly used their professional judgement to complete an OASys assessment. Seven out of ten felt that the amounts of professional judgement required to complete an assessment were about right, but approximately one quarter stated it was too little.
- There were some clear resource issues, with over half of the assessors disagreeing that they usually had sufficient working hours to complete assessments. More than four fifths of the assessors felt there was scope for improving the targeting of OASys and its layers.
- Approximately two thirds of the assessors felt that the level of quality assurance for OASys assessments was about right, and over half thought that the introduction of the quality assurance process had raised the quality of OASys assessments. However, around three quarters believed that there remained scope for improving quality.

Areas where awareness needed to be raised included: (i) the value and workings of the actuarial reoffending predictors; (ii) the current targeting criteria for the OASys layers and fast reviews; and (iii) the availability of the online help.

2.1 Context

Currently, there are an estimated 12,000 OASys assessors. Previous studies obtaining the views of OASys users had small samples and produced findings relevant to OASys in early stages of roll-out, and prior to prison/probation connectivity and other more recent developments.[3] The latter include the

[3] Gloucestershire Probation Area (2004); Cornwell Management Consultants (2005); Mair, G., Burke, L. and Taylor, S. (2006). A study has recently been conducted with YOT practitioners to gather their perceptions of Asset, the risk assessment tool for young offenders (Wilson and Hinks, 2011).

changes implemented in August 2009 which streamlined the original (full) assessment while introducing a new shorter (standard) assessment (as well as a basic assessment for use in the community) and a fast review facility. Bearing in mind the importance of user acceptability and the need for OASys to support everyday practice as well as possible, it was recognised that these latest developments would benefit from a robust form of user feedback.[4]

2.2 Approach

Research questions

The overall aim of the study was to capture prison and probation assessors' views and experiences of OASys, highlighting further potential improvements to the tool and associated processes. It was deemed vital that both prison and probation assessors were included, recognising that the differing targeting of OASys alongside the differing regimes, functions and cultures of the two services could have an impact on their views regarding OASys.

To guide the research, the following six key objectives were set:

1. To identify the perceived strengths and weaknesses of the OASys assessment process.
2. To identify the perceived strengths and weaknesses of the targeting of OASys.
3. To identify the perceived strengths and weaknesses of the content of OASys.
4. To identify the perceived strengths and weaknesses of the OASys training and guidance.
5. To identify the perceived strengths and weaknesses of the OASys quality assurance procedures.
6. To identify potential improvements to be taken forward following the implementation of the new OASys IT system (delivered through the OASys replacement (OASys-R) project).[5]

The study was conducted in two parts. Firstly, assessors' perspectives were explored through an online self-completion questionnaire (Dillman, Smyth and Christian, 2009).[6] Secondly, a small number of individual face-to-face structured interviews were conducted with OASys leads/managers. At both stages, the questions were framed around the themes of the study's objectives:

[4] Feedback is currently non-systematic and largely anecdotal, being made via OASys leads in prisons and probation trusts to the OASys business team.

[5] As improvements to the IT system were being taken forward through the OASys-R project, the focus was upon non-IT issues. However, the online questionnaire did include a question asking assessors whether IT issues had a detrimental impact upon their ability to complete an OASys assessment – nearly two-thirds (64%) said that they did.

[6] A web-based survey was developed and accessed through 'SurveyMonkey' (http://www.surveymonkey.com). SurveyMonkey had previously been used successfully by NOMS colleagues across the probation and prison IT systems. Its standard survey designs are fully compliant for respondents with disabilities.

Objective 1: Assessment process

- How does the information collected in OASys support practitioners in managing offenders' risks and needs? Is this support sufficient?
- How is professional judgement used or exercised in completion of OASys? Does OASys allow for an appropriate amount of professional judgement?
- How are offenders engaged in the assessment process? Is this sufficient?
- How is the self-assessment questionnaire used to inform the core assessment and sentence plan?
- How are the different predictors of reoffending (OGP1, OVP1, OGRS3) being applied?
- How well does OASys assist with completing reports (e.g. Pre-Sentence Reports)?
- How well does OASys link to other more specialist assessments?
- How well does OASys assist with the targeting of interventions?
- How well does OASys enable change and progress to be monitored?
- What are the strengths and weaknesses of layered OASys (compared to the pre-layered system)?
- Is sufficient time available to complete OASys?

Objective 2: OASys targeting

- Are the OASys layers targeted at the most appropriate offenders?
- Are OASys fast reviews used in appropriate cases?
- How has layered OASys helped in terms of resource demands?
- How have fast reviews helped in terms of resource demands?

Objective 3: OASys content

- What information is available for completing OASys and what information is needed?
- What are the general strengths and weaknesses of each of the OASys components (the core assessment, the RoSH assessment, the offender self-assessment questionnaire and the sentence plan)?
- Is the right level of detail collected at the basic, standard and full layers?
- Is the right level of detail collected in fast reviews?
- Are any important risk factors missed? What additional risk factors are suggested for inclusion / consideration?
- Are any important protective factors missed? What protective factors are suggested for inclusion / consideration?
- Does OASys record any unnecessary information?
- How well does OASys cater for all offender groups?
- How well does OASys deal with diversity issues?

Objective 4: OASys guidance and training

- Is OASys training available when required?[7]
- How well does the training meet users' needs?
- What are the strengths and weaknesses of joint prison/probation training?
- Is the OASys guidance manual and online help sufficiently clear and detailed?

Objective 5: OASys quality assurance processes

- Are quality assurance procedures sufficient?
- What are the strengths and weaknesses of joint prison/probation quality assurance procedures?
- Do users feel that the quality of assessments is high?
- Do users feel that the introduction of the OASys QA tool and its associated processes has helped to raise the quality of assessments?
- Do users feel that there is consistency between different assessors?
- Do users feel that there is consistency between prison and probation assessments?

Objective 5: Potential improvements

- What improvements could be made to the targeting of the OASys layers?
- What improvements could be made to the content and structure of OASys?
- What improvements could be made to OASys training and guidance?
- What improvements could be made to quality assurance procedures?

Both the questionnaire and the interview schedule were piloted with a small number of users to assess whether: (i) they could be completed within reasonable timeframes; (ii) the questions were clear and understandable; and (iii) any specific questions needed to be amended, added or omitted. It was checked that the online questionnaire operated correctly across the probation and prison IT systems and that it was sufficiently respondent-friendly.[8] When conducting the interviews, attention was given to ensuring that the interview questions were fully understood, with clarification provided when necessary. An interview guide was used to ensure the same key questions were addressed.

[7] For the prison service, training is managed centrally; whereas for the probation service, training is managed by the individual trusts.

[8] Users complete OASys assessments in the prison and probation IT systems, and are thus sufficiently IT literate. A Welsh language version of the survey was not developed, taking into account time and resource constraints, and the fact that the OASys IT systems and OASys communications are in English only.

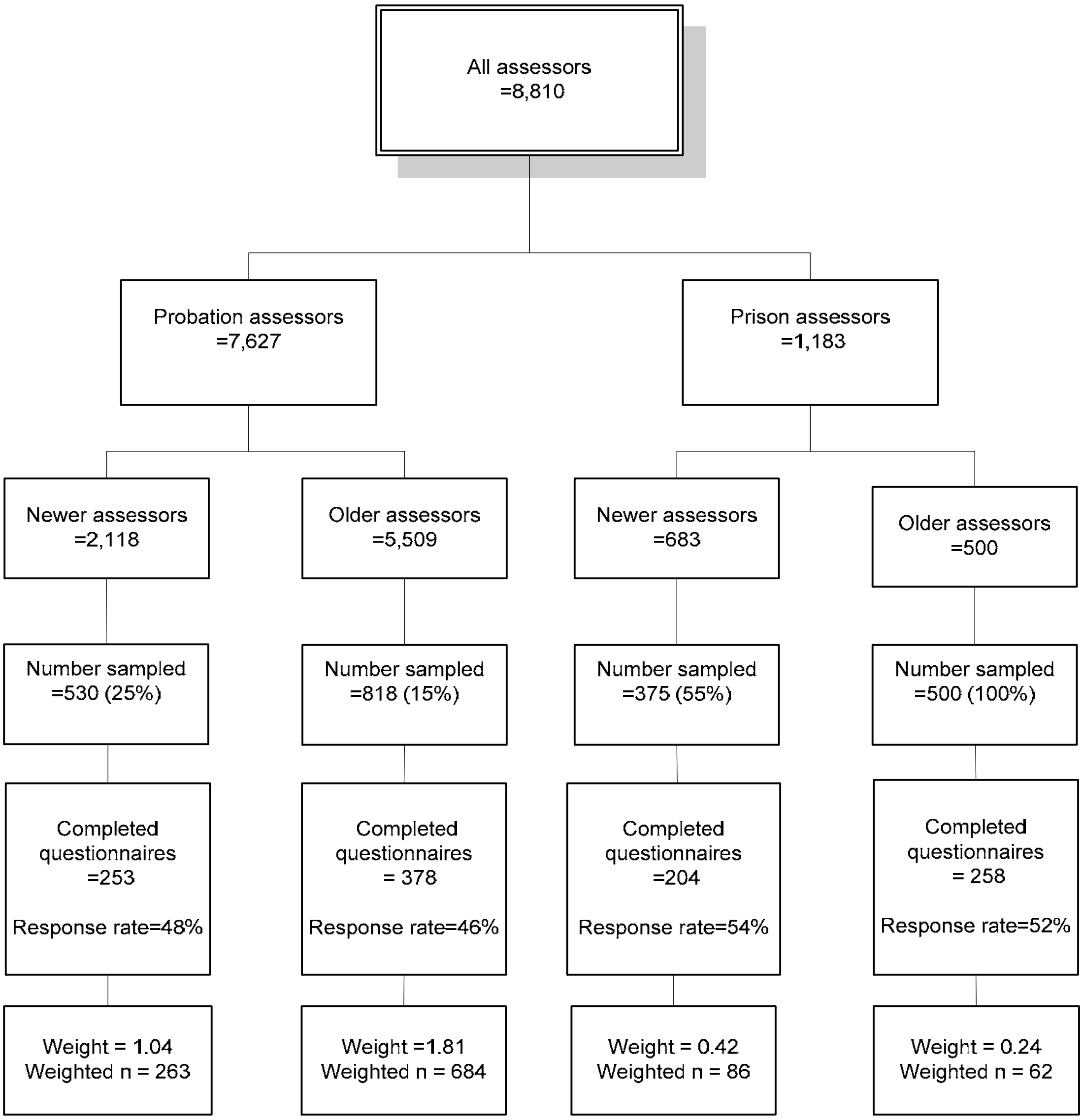
The samples

The use of the online questionnaire enabled a broad cross section of OASys assessors to be targeted, with a disproportionate stratified random sample being extracted from the O-DEAT database. The assessments held within this database included the names of the OASys assessors who have completed the assessments. It was thus used to create a sampling frame of both prison service and probation service assessors. De-duplication ensured that an assessor was only included once on the sampling frame list. During the first quarter of the financial year 2010/11 (April to June 2010), assessments were completed by 8,810 different users across the 35 probation trusts and 133 prisons. Of these 8,810 users, the majority were from the probation service: 7,627 (87%) assessors compared to 1,183 (13%) prison service assessors. 2,801 (32%) of the assessors could be considered 'new' assessors, having completed their first assessment no earlier than 2007/08. But the proportion of 'new' assessors differed greatly between the two services; 58% in the prison service compared to 28% in the probation service.

As shown by Figure 2.1 below, the sample was disproportionately stratified (Sapsford, 1999) to ensure that it was representative of probation and prison assessors and of those who were relatively new to OASys and those who had been using the tool for sometime. To ensure sufficiently large samples for all four strata were obtained, the following sampling proportions were used:

- New probation assessors: 25% – 530 / 2,118
- Older probation assessors: 15% – 818 / 5,509
- New prison assessors: 55% – 375 / 683
- Older prison assessors: 100% – 500 / 500

Figure 2.1: Sampling approach for OASys user perspective survey



In total 2,223 assessors were contacted at the start of April 2011 The email addresses for each of the prison and probation assessors in the sampling frame were obtained and verified through the internal NOMS email system, the Probation Directory (National Association of Probation Officers, 2011) and the Electronic Probation Information Centre (EPIC) – the Probation intranet. One disadvantage with using the O-DEAT database as the sampling frame was that in some circumstances, it was not possible to confirm whether the individuals were still working as an OASys assessor, or even whether they were still working within a prison establishment or probation trust.[9] When an email delivery error message was received to say the email address was undeliverable, efforts were made to obtain the correct email address.[10]

[9] In some instances, individuals responded to emails asking them to participate in the survey to say they were no longer OASys assessors; these people were removed from any further communication or follow-up emails regarding the survey.
[10] Replacement assessors were not selected when individuals could not be contacted.

A response rate of around 50% was anticipated,[11] resulting in an acceptable margin of error for the total sample – a maximum 2.7% margin of error[12] (50/50 response rates) at the 95% confidence level for a final sample of 1,111 assessors.[13] Early communication and close liaison with both services ensured that the purpose of the research was clearly understood, while encouraging sufficient time to be allowed for staff participation. While assessors were not offered incentives for their participation, they were made aware that their contribution would enhance the further development and improvement of OASys. The use of the online survey enabled quick monitoring of survey progress and allowed multiple reminders to be issued easily. Following the issuing of these reminders, a final sample size of 1,093 assessors was achieved, representing a response rate of 49%. Potential unit non-response bias, due to distinct differences between those who responded and those who did not, was checked by comparing the two groups in terms of their locations and OASys experience (Berg, 2005). To correct for the differing sampling ratios and response rates across the four strata, a non-inflationary, proportional design weight was applied,[14] calculated as % of stratum in the population / % of stratum in the final sample. Thus, for old probation assessors, the weight was (5,509 / 8,810) / (378 / 1093) = 1.8.

Item non-response was also checked by comparing the response rates to individual questions across the four different sample groups. Having examined the patterns of missing data,[15] the decision was taken to impute values to replace any missing values. Multiple imputation was used, imputing the data five times over and then pooling the results (Wayman, 2003).[16] For each question, key assessor details (prison or probation, probation grade, age, gender and ethnicity) and responses to all the other questions were entered into the models as predictors.

At the second stage of the study, twelve face-to-face interviews were conducted with OASys leads/managers, examining the respondents' more strategic perspectives and their knowledge of the OASys completion process and risk assessment practice more generally. For this stage of the study, non-representative convenience sampling was used; the purpose being to explore further, through open-ended questions, the issues raised in the online survey. The OASys leads/mangers were identified through the OASys business team who asked for volunteers to participate. It was

11 Response rates for two previous online surveys conducted by NOMS colleagues were 53% (video conferencing project; 111/209) and 67% (Layered Offender Management project; 65/97).

12 The margin of error is a common summary of sampling error that quantifies uncertainty about a survey result. For example a margin of error of 2.7% indicates that if, for example, 45% of the sample respond in a particular way we can be confident that if the entire population were asked the same question between 42% and 48% would also answer in the same way.

13 This margin of error is for the whole sample and does not take into account the stratification that was used.

14 Non-inflationary, proportional weights make the weights sum to the sample total (rather than grossing-up to the population total) through inflating the under-sampled cases and deflating the over-sampled ones. Correcting the proportions in the sample without increasing the scale of the figures is advisable when conducting significance tests in SPSS Statistics.

15 Whilst the levels of missing data were low at the beginning of the questionnaire i.e. around 1% of all cases, the levels were higher towards the end of the questionnaire, i.e. approximately 15% to 20% of all cases.

16 An iterative Markov chain Monte Carlo (MCMC) method was used. For most questions, the results when including and excluding the imputated values for the missing cases did not differ greatly – the differences were usually less than five percentage points. The main exception was Q39 (Overall how would you personally rate the quality of OASys assessments in other trusts/establishments?) - the results across the three-point scale differed by sixteen, nine and seven percentage points.

nevertheless ensured that differing types of prison establishments and differing sized probation trusts were represented.

Analysis

The combination of fixed response and open-ended questions within the online survey enabled (i) comparisons to be easily made between different users and between the prison and probation services while (ii) allowing assessors to provide further explanation and elaboration. The responses to all fixed-response questions were analysed through use of IBM SPSS Statistics, assessing whether certain views were related to the assessors' location (prison or probation), grade, OASys experience, gender, age or ethnicity,[17] or were more typical of general views and responses. Chi-square tests were used to assess whether the views between assessor sub-groups differed significantly.[18]

The responses to the open-ended questions in the self-completion questionnaire were explored via the text analytics component of IBM SPSS Modeller. This text mining tool employs advanced linguistic technologies and Natural Language Processing to extract key concepts/terms.[19] For the open-ended questions included within the face-to-face interviews, a thematic analytical framework was developed with links to the research questions.

Limitations

Some potential non-response biases from use of the online survey were checked (see above), but other potential biases remained. For example, enthusiastic users or users with a more negative attitude to OASys might have been more likely to respond than indifferent users.

As noted above, no attempt was made to ensure that the second stage interviews were fully representative. Furthermore, only a small number of interviews were conducted at this stage meaning any results will not be generalisable to the wider OASys lead or manager population. In order to save time and resources, these interviews were conducted through the use of video conferencing technology (VCT). However, in some circumstances it was not possible to use the VCT facilities, with one interview being conducted face-to-face and two being conducted over the telephone. It is possible that these differing forms of engagement may have affected the responses obtained (Oppenheim, 1999).

17 All statistically significant differences at the 95% confidence level were examined, but this chapter focuses upon such differences between (i) prison and probation assessors and (ii) Probation Officers (POs) and Probation Service Officers (PSOs).

18 In relation to probation grades, chi square tests were only performed between PSO and PO grades due to low base numbers for Senior Probation Officer (SPO) grades. Similarly, chi square tests were not performed on ethnicity due to low base numbers for Black, Asian and Minority Ethnic (BME) assessors.

19 Key concepts/terms, representing the essential information, were extracted automatically, with normalisation and grouping techniques correcting punctuation and spelling errors respectively. Closely related concepts were then grouped into higher-level categories, firstly through further linguistic based methods, identifying synonym and hyponym relationships and root terms, and then manually.

The results of the linguistic text mining, used to extract the textual information recorded within the completed online questionnaires, are dependant upon the linguistic resources used. The dictionary resources include synonyms, words to be excluded from extraction, types that group together multiple terms, and other more specialised tuning algorithms, such as words not to be confused when fixing spelling errors. Further editing of these resources through multiple iterations could improve the accuracy and value of the concepts extracted.

2.3 Results

The assessment process

Approximately nine out of ten assessors (89%) felt that the information recorded in an OASys assessment supported them very or fairly well in managing offenders' risks and needs. Assessors were specifically asked about the usefulness of the differing risk predictors/ratings in managing offenders' risks and needs. As shown by Table 2.1, the RoSH ratings, which are based upon structured professional judgement, were most commonly viewed as very or fairly useful – 90% of assessors responding in this way. For the actuarial reoffending predictors (OGP1, OVP1 and OGRS3), the proportion of assessors viewing them as very or fairly useful were 67% for OGP1, 68% for OVP1 and 70% for OGRS3.[20] Across all three actuarial predictors, there were significant differences between the views of Probation Service Officers (PSOs) and Probation Officers (POs), with the former more likely to view the predictors as very or fairly useful.[21] For example, three-quarters (75%) of PSOs viewed OGP1 as very or fairly useful compared to less than two-thirds (64%) of POs, with 16% of the latter responding that it was not very or not at all useful.

Within the interviews, the majority of the OASys leads tended to agree that the RoSH ratings were most useful, with the use of professional judgement underpinning these ratings being highlighted as a positive. Other explanations for the differing views regarding the predictors/ratings included a preference for textual summaries and low, medium and high ratings rather than a quantitative score or percentage figure.

> *"In terms of accessibility, when you want to quickly look something up and get a quick overview of somebody you are about to interview who you don't know, I would look for the words rather than numbers."* Probation, Quality development officer

Another interviewee felt that the RoSH ratings were easier to understand than scores, and admitted that they did not have a good understanding of the actuarial scores or percentages. There also appeared to be some confusion about the actuarial scoring. A couple of interviewees were not aware

20 All these actuarial reoffending predictors have been validated as robust measures (Howard, 2009; Howard *et al.*, 2009). Differences in assessors' views between the usefulness and application of actuarial and clinical measurements has also been highlighted through previous research with probation officers (Fitzgibbon *et al.*, (2010), Robinson, (2002)).

21 A Canadian study, based on 71 interviews with correctional workers, reported that some respondents felt that actuarial assessments were most useful for less experienced staff, with seasoned professionals being able to use their "experiential knowledge" and/or "common sense" (Hannah-Moffat, Maurutto and Turnbull (2009)).

that the actuarial reoffending predictors also produced low, medium and high ratings or that they incorporated both static and dynamic factors.

> *"To me I look at the stats and say it's a high score, but for me I can't internalise that, whereas if I could think in terms of low, medium, high that's better for me, it's more tangible."* Prison, Offender supervisor

> *"RoSH ratings are easier compared to scores – [some assessors] don't know what the scores mean. It is a training/knowledge issue and use of language."*
> Prison, Head of Offender Management

Table 2.1: Assessors' views on the OASys assessment process

Question	Unweighted n	Answer (%)		
How well do you feel:		**Very / fairly well**	**Neither well or not well**	**Not very / not at all well**
• the information recorded in an OASys assessment supports you in managing offenders' risks and needs?	1,093	89	8	3
• that an OASys assessment assists with the targeting of accredited interventions?	1,093	57	25	19
• that an OASys assessment assists with the targeting of non-accredited interventions?	1,093	54	28	18
• that specialist assessments (e.g. RM2000, SARA) link to OASys assessments?	1,093	38	28	34
• OASys assessments assist with completing standard delivery Pre-Sentence Reports?	631*	67	13	20
How useful do you feel the following risk predictors/ ratings are in managing offenders' risks and needs?		**Very / fairly useful**	**Neither useful or not useful**	**Not very / not at all useful**
• OGP1	1,093	67	21	13
• OVP1	1,093	68	20	12
• OGRS3	1,093	70	17	13
• Risk of Serious Harm ratings	1,093	90	6	4
To what extent do you agree or disagree with the following statements?		**Strongly agree/ agree**	**Neither agree nor disagree**	**Disagree / strongly disagree**
• OASys assessments enable progress in addressing offenders' risks and needs to be monitored.	1,093	80	13	7
• OASys assessments enable changes in offenders' risks and needs to be monitored.	1,093	81	12	7
• I usually have sufficient working hours to complete OASys assessments.	1,093	29	15	56
• I regularly use my professional judgement to complete an OASys assessment.	1,093	93	4	3

Question	Unweighted n	Answer (%)		
Generally, I feel that the amount of:		About right	Too little	Too much
• professional judgement required to complete an OASys assessment is…	1,093	71	24	5
• offender engagement in the sentence planning process is…	1,093	69	28	3
		Yes	No	
Generally, do you feel that you have sufficient offender information to complete an OASys assessment?	1,093	68	32	-
Do you use the offender self-assessment questionnaire to inform the sentence plan?	1,093	84	16	-
Do you feel that IT issues have any detrimental impact upon your ability to complete an OASys assessment?	1,093	64	36	-

*indicates a probation only question

Professional judgement

The majority (93%) of the assessors strongly agreed or agreed that they regularly used their professional judgement to complete an OASys assessment (Table 2.1).[22] A higher percentage of probation assessors strongly agreed or agreed they regularly used their professional judgement compared to prison assessors (94% compared to 89%). Differences were also noted when the assessors were asked about the amounts of professional judgement they felt were required to complete an OASys assessment. Overall, approximately seven out of ten (71%) of the assessors felt that the amounts were about right, but approximately a quarter (24%) stated it was too little. Probation assessors were more likely than prison assessors to respond that too little professional judgement was required to complete an assessment (27% compared to 11%), with just under a third (32%) of POs responding in this way compared to 15% of PSOs.

During the OASys lead interviews, mixed views were expressed on the amounts of professional judgement required and whether more should be used. There were suggestions that the structure of the OASys assessment could, in some instances, make it feel to the assessor that they were not using their professional judgement.

> *"I think a lot of professional judgement is required in OASys but sometimes the way things are worded and some of the tick boxy bits of OASys make it feel like professional judgement isn't being exercised."* Probation, OASys countersigner and quality assurer

[22] Similarly, in a Canadian study, correctional workers reported that professional judgment was "regularly incorporated into the actuarial assessment process" (Hannah-Moffat, Maurutto and Turnbull, 2009). The practitioners continued to "exercise considerable judgment in determining the selection of information, the identification of collateral sources to be consulted, the assessment of criteria, the calculation of risk scores, and the extent to which risk outputs determine recommendations and case management".

It was noted that the text boxes within the assessment could allow an assessor to expand upon and use their professional judgement. If more professional judgement were to be used, the interviewees noted that it would be important to ensure that:

(i) assessors had sufficient confidence in their own abilities; and
(ii) sufficient time was made available to enable them to complete assessments which were of sufficient quality.

"It's also about confidence, not just about not having the option"

Probation, OASys quality assurer

Two of the interviewees who had been involved in 'Professional Judgement pilots' noted how there had been a shift in attitudes amongst assessors towards OASys. The pilots had relaxed the probation National Standards on the timescales for OASys completion.

"Certainly in [probation area]…[there has] been a shift through the pilot from people thinking OASys is a very lengthy form they have to fill in before they can get on with doing their job, towards thinking OASys is there to serve us and we need to be using our judgements and assessments, and we need to make a decision about how OASys is going to best help us record those things. In a sense, OASys should be all about professional judgement and I think it's about trying to help practitioners see it in a different way to make it more about professional judgement."

Probation, OASys countersigner and quality assurer

Offender information for the OASys assessment

Approximately two-thirds (68%) of the OASys assessors felt they had sufficient offender information to complete an OASys assessment (Table 2.1). The views of prison and probation assessors differed significantly, with a higher proportion of prison assessors compared to probation assessors stating they did not have sufficient offender information (50% and 29% respectively). Where the assessors stated the information was insufficient, the reasons included difficulties gaining access to:

(i) earlier reports such as Pre-Sentence Reports;
(ii) Crown Prosecution Service documents including the disclosure pack;
(iii) information from previous establishments; and/or
(iv) previous offence history for all types of offenders, and, more specifically, non-UK offence information for Foreign National Prisoners.

It was evident from the OASys lead interviews that there were local differences in the ease with which relevant information could be obtained. Areas gave examples of local initiatives or changes in the ways of working which had helped overcome some problems. These included (i) the merging of court administrative teams and a more consistent presence in court, leading to better access to key reports, and (ii) establishing direct contact with partner agencies.

Offender engagement in the sentence planning process

Approximately seven out of ten (69%) assessors felt that the amount of offender engagement in the sentence planning process was about right; with over a quarter (28%) saying it was too little. Significant differences were noted between prison and probation assessors with a higher proportion of probation assessors stating it was too little (30% and 18% respectively).

While most interviewees thought that there was scope for increasing the levels of offender engagement, having the time to fully engage with an offender was seen as a key issue.

> *"By and large we have as much as we can, because we really are busy. A lot of it is very reactive. We often don't get the time to work one-to-one with somebody and build up a relationship with them, where you feel like you can make a difference in offending behaviour. Biggest complaint – everybody would like to be able to spend more time. Considering the constraints and the amount we have to juggle we do a good job, but we would like to do more."* Prison, Offender supervisor

The majority (84%) of the OASys assessors responded that they used the offender self assessment questionnaire (SAQ) to inform the sentence plan. Respondents stated that the SAQ helped to identify problem areas or areas of need for the offender. Others highlighted the importance of gaining the offenders' opinions and an understanding of their perspective, which could then be incorporated within the sentence planning process. A higher percentage of probation assessors compared to prison assessors said they used the offender SAQ to inform the sentence plan (86% compared to 78%). Furthermore, a higher percentage of PSOs than POs used the offender SAQ in this way (92% and 81% respectively).

For those who responded that they did not use the SAQ to inform the sentence plan, this was often because they were not involved in the sentence planning process. Some had a preference for using information from (i) the core OASys assessment and/or (ii) face-to-face discussions with offenders. Others felt that the form was too basic and did not provide sufficiently reliable information.

> *"More of a 50/50 [in terms of SAQ use] rather than an outright 'no'. Sometimes a good SAQ marries up well with the assessor's own observations, other times the offender can have very limited insight or wishes to conceal some issues and therefore the SAQ is not completed with insight or integrity."* Probation assessor (Questionnaire response)

During the interviews, the OASys leads highlighted scope for improving the offender SAQ form. One probation interviewee felt that, whilst more people were using the SAQ, it was not very 'solution focused' because of the yes/no format of answers to set questions. The position of the offender SAQ, which was placed towards the end of the OASys operational IT systems, was also seen by some OASys leads as unhelpful.[23]

OASys resource demands

Over half (56%) of the assessors strongly disagreed or disagreed with the statement 'I usually have sufficient working hours to complete OASys assessments' (Table 2.1). There were differences between probation grades, with a higher percentage of POs compared to PSOs strongly disagreeing or disagreeing that they usually had sufficient time (63% and 48% respectively). A higher percentage of probation assessors compared to prison assessors also strongly disagreed or disagreed they usually had sufficient time (57% and 47% respectively).

During the OASys lead interviews, mixed feelings were expressed regarding the amount of time assessors had to complete OASys assessments. Whilst some interviewees felt there was an issue with the amount of time required for completing good quality assessments, others disagreed:

> *"Yes, most people (across all grades POs PSOs) feel that they don't have sufficient time to complete an assessment to the standard that they would like. It's taking time away from other things that they would rather be doing."*
>
> Probation, OASys countersigner and quality assurer

> *"Possibly not now. Originally when [OASys] first rolled out it was [an issue with the amount of time required to complete a quality assessment]. But I think that the assessors we have are now all up to speed, apart from two new ones."* Prison, Transfer Officer

Nine out of ten (90%) of the assessors were OASys users prior to the changes made to OASys in August 2009. Of these, more than half (53%) felt that the introduction of the standard OASys layer was very or fairly helpful in terms of reducing the time they spent completing assessments (Table 2.2). The views of the POs and PSOs differed significantly – 71% of PSOs found the introduction of the standard OASys layer helpful, compared to less than half (45%) of POs. This difference could be due to the PSOs handling a larger proportion of the lower risk cases and thus using standard assessments more regularly.

---

[23] The position of the SAQ was changed as part of the OASys-R project. It now appears before the RoSH component.

Table 2.2: Assessors' views on the August 2009 changes to OASys

Question	Unweighted n	Answer (%)			
		Yes	**No**		
Were you an OASys user prior to the changes made within OASys release 4.3.1 (August 2009)?	892	90	10		
How helpful do you feel the introduction of the following have been in terms of reducing the time you spend completing assessments?		**Very/fairly helpful**	**Neither helpful or unhelpful**	**Not very/not at all helpful**	**Don't know**
• Standard OASys layer	780	53	27	15	6
• Fast reviews	780	58	15	17	9
Do you think the basic OASys layer has been a helpful addition?	457*	46	20	17	18

*indicates a probation only question

As Table 2.2 also shows, just under half (46%) of the probation assessors who used OASys prior to the August 2009 changes viewed the introduction of the basic layer as a very or fairly helpful addition.[24] During the OASys lead interviews, mixed feelings were expressed. Whilst some felt that the information collected was too simplistic, others thought the detail was adequate. Just under three-fifths (58%) of all the pre-August 2009 assessors felt that the introduction of fast reviews was very or fairly helpful in terms of reducing the time they spent completing assessments. There was a significant difference between the views of prison and probation assessors; approximately three-fifths (61%) of probation assessors noted a positive impact of fast reviews on time spent completing assessments compared to approximately two-fifths (39%) of their prison counterparts. This difference may be attributed to the fact that assessments are reviewed less regularly within a prison setting, limiting the appropriateness of fast reviews. Within the probation service, 72% of PSOs stated that fast reviews had saved time, compared to 55% of POs.

Whilst some OASys leads felt that the basic layer and fast reviews reduced the time spent completing assessments for lower risk cases and increased their face-to-face time with higher risk offenders, others thought that fast reviews did not save as much time as they could, pointing out that they could not be used whenever the questions in a specific section required updating.

24 While basic assessments have been completed by all probation trusts, their initial uptake varied hugely. Thus, while the question on basic assessments was asked of all probation assessors who were users prior to August 2009, some assessors will have been able to reach a more informed judgment than others.

Targeting of the OASys assessment

More than four-fifths (83%) of all assessors felt there was scope for improving the targeting of OASys and its layers (Table 2.3), with probation assessors significantly more likely to agree with this view compared to prison assessors (84% and 78% respectively). Of all assessors who held this view, around a quarter (24%) felt the full layer was targeted at too many offenders, while one-fifth (20%) believed that fast reviews were targeted at too few offenders.

> *"I agree fast reviews are used with too few offenders. I think that's about potentially changes in practice rather than targeting. I think the targeting is okay but because people have the option to do what they've always done it's easier to keep doing that rather than learn to do something differently."* Probation, Quality Development Officer

Table 2.3: Assessors' views on the targeting of the OASys assessment

Question	Unweighted n	Answer (%)			
		Yes	**No**	**-**	**-**
Do you believe that there is scope for improving how OASys is targeted at offenders?	1,093	83	17		
Do you feel that the following types of OASys are targeted at the most appropriate offenders?		**Yes, about right**	**No, too many offenders**	**No, too few offenders**	**No, other**
• Basic	531*	64	15	14	7
• Standard	888	66	16	10	8
• Full	888	63	24	8	5
• Fast reviews	888	57	12	20	11

*indicates a probation only question

Some assessors, and also OASys leads, indicated that a decision had been taken in their trust/establishment not to complete a specific type of assessment or fast reviews. Fast reviews were not considered to be as resource friendly as had been initially suggested.

> *"[There is] confusion over fast reviews. I'm not sure how fast they are. We've got to do the work either way. We have to go through the work even if we are not sure whether it should be a fast review or not. There is a push for assessors to do more fast reviews wherever possible because it's supposed to be resource friendly but the outcome could be different with assessors having to go back and do another one (assessment)."*
>
> Probation, Practice Manager

Furthermore, some OASys leads perceived that the different layers and fast reviews were not always used correctly because of a lack of knowledge and/or training.

> *"I would agree. I'm confused. I'm clearer now about the different layers. What I need to get my head around is the fast reviews. And that's because the team that I work in largely deal with Tier 4 offenders, so we haven't had the option other than a full, standard review. And now that is a shift in practice. Mainly doing full and standard and not fast reviews. I think there is scope to do fast reviews but we are not fully aware of what scope there is."*
>
> Probation, OASys countersigner and quality assurer

Overall, the majority of interviewees agreed that there was scope for improving the targeting of the different layers. When OASys leads were asked to suggest what could be done to improve the knowledge of how to use the different OASys layers and fast reviews, suggestions included:

- Team managers helping to ensure that the correct layer is used and, if necessary, having discussions with assessors when the wrong layer of assessment has been used.
- Creating a discussion database where assessors can access relevant information on the different layers and their use.
- Asking staff members to assess the quality of other assessors' work.
- Sending out local guidance and practice instructions.
- Staff being given the opportunity to attend briefings which provide further information on the OASys layers, rather than only being sent information about OASys changes via email correspondence.

> *"For me personally the guidance has been in a written form and I have had or made the time to sit down and get my head around it. It's not my preferred way of receiving information. I would rather somebody sat down and talked to me about it and perhaps even went through one with me. So I think it is probably implementing changes via email."*
>
> Probation, OASys countersigner and quality assurer

"I think the difficulty is the people writing the guidance know what they are doing, so they write the guidance from a point of having a great deal of knowledge which sometimes isn't helpful guidance for the person who has no knowledge."

Probation, Quality Development Officer

Content of OASys assessments

Level of detail and content

When assessors were asked about the level of detail and content in the different types of OASys, the proportion of assessors viewing them as about right was 63% for the basic (probation assessors only), 68% for the standard, 73% for the full and 68% for fast reviews (Table 2.4). More than one in ten felt that too much detail was collected in the standard and full layers of OASys (13% for both respectively), but a similar proportion (12%) felt that too little detail was collected at the standard layer. Some highlighted the difficulties that could be caused by switching between the two layers, with some valuable information not being pulled through to subsequent assessments.

Table 2.4: Assessors' views on the content of the OASys assessment

Question	Unweighted n	Answer (%)			
Do you feel that the right level of detail is collected in the following types of OASys?		**Yes, about right**	**No, too much**	**No, too little**	**No, other**
• Basic	631*	63	9	19	9
• Standard	1,093	68	13	12	7
• Full	1,093	73	13	7	7
• Fast reviews	1,093	68	8	16	8
To what extent do you agree or disagree with the following statements?		**Strongly agree/agree**	**Neither agree nor disagree**	**Strongly disagree/ disagree**	
I feel that the content of OASys is sufficient for assessing all offenders.	1,093	59	19	22	-
How well do you think the OASys assessment deals with the following diversity issues?		**Very/fairly well**	**Neither well or not well**	**Not very/not at all well**	
• Age	1,093	49	29	22	-
• Gender	1,093	49	28	23	-
• Ethnicity	1,093	49	26	25	-
• Disability	1,093	52	25	23	-
• Religion	1,093	46	31	22	-
• Sexual orientation	1,093	36	35	29	-
		Yes	**No**	**Don't know**	
Do you feel that any unnecessary information is required to be recorded at any of the OASys layers?	1,093	29	71	-	
Do you feel that any risk factors are missed from all of the OASys layers?	1,093	21	59	20	
Do you feel that any positive/protective factors are missed from all of the OASys layers?	1,093	16	61	23	

*indicates a probation only question

The proportion of assessors stating that too little detail was collected at the basic layer and within fast reviews was 19% and 16% respectively (Table 2.4). As Table 2.4 also shows, approximately three-fifths (59%) of assessors agreed that the content of OASys was sufficient for assessing all offenders.

Comparing probation grades, PSOs were more likely to agree with this statement than POs (68% and 59% respectively).

Information collected in an OASys assessment

Approximately seven out of ten (71%) of the assessors did not feel that unnecessary information was required to be recorded at any of the OASys layers (Table 2.4). Probation assessors were more likely to feel unnecessary information was required than their prison counterparts (31% versus 14%). Similarly, POs were more likely to state this view compared to PSOs (33% versus 20%).

Some assessors made specific reference to overlaps between the summary and full RoSH sections. OASys leads differed in their views; some agreed that the RoSH sections were 'long-winded' with a lot of information either being carried over or just 'cut and paste' from the different sections. But others stated that the summary sheet was useful, providing a 'quick snapshot of all issues and scores'.

Some assessors also made specific reference to overlaps between the sentence plan and risk management plan, with which the majority of OASys leads agreed. For example, details on professional contact and agency involvement in the risk management plan were often duplicated in the liaison arrangement of the sentence plan.

When asked about duplication of questions within the core OASys assessment, prison and probation OASys leads felt that there was extensive repetition in the information being recorded, notably within the evidence boxes at the end of each section. However, one interviewee felt that some degree of repetition was inevitable, as offenders' risks and need factors were intertwined and could not be seen in isolation.

Interviewees agreed that duplication occurred in the following sections:

(i) Section 2 (analysis of the offence) – this section was considered too repetitive; only the pertinent information should be pulled through.

(ii) Section 9 (alcohol misuse) – 9.1 two text boxes – Information from the Q9.1 text box (‘If a problem describe level and frequency of alcohol consumption at present time’) is repeated in the evidence ratings text box at the end of the section.

(iii) Sections 11 (thinking and behaviour) and 12 (attitudes) – these two sections were considered to be 'quite closely linked and could be incorporated into one'.

Risk factors

Approximately three-fifths (59%) of assessors thought that no risk factors were missing from the OASys layers, with approximately one-fifth (21%) disagreeing (Table 2.4). Probation assessors were more likely to indicate that risk factors were missing compared to prison assessors (23% and 12% respectively), with POs more likely to hold this opinion than PSOs (30% and 10% respectively).

Some assessors who said that risk factors were missing did not identify specific factors, but indicated that OASys was too general for specific offender groups such as female sex offenders, female domestic violence perpetrators, internet sex offenders and serving or ex-serving military personnel. Others, on the other hand, cited the following risk factors they felt could be included or expanded upon within the assessment:

(i) Domestic violence (DV): jealousy; number of DV call outs; number of police call outs/police intelligence; any presence of children in the household; prostitution; historic DV history; vulnerability to harm posed by others; power/controlling behaviours; any current restrictions e.g. harassment order.

(ii) Sex offenders: sexual interest; sexual experiences; issues relating to sexuality.

(iii) Mental health: nature of delusional beliefs/hallucinations; psychosis; personality disorder; Asperger's syndrome/autism; Attention Deficit Hyperactivity Disorder (ADHD).

(iv) Other: gambling; gang-related issues.

OASys leads had mixed views when they were asked whether specific risk factors relating to DV and sex offenders should be included. In terms of sex offenders, some interviewees felt that OASys was not well tailored for this group or particular sub-groups (e.g. child sex offenders) or for highlighting sexual offences which were not the index offence. Some interviewees thought that specific prompts should be made to Risk Matrix 2000 (RM2000).[25] On some occasions, OASys leads suggested that assessors would complete RM2000 on paper but this information was not reflected in the OASys assessment. Furthermore, one third of assessors felt that OASys assessments did not link very or at all well with specialist risk assessments such as RM2000 or Spousal Assault Risk Assessment (SARA)[26] (Table 2.1).

In terms of DV, some OASys leads felt that this area was sufficiently covered within the relationships section (which includes a textual box for recording relevant information) and within SARA. Others, on the other hand, felt that DV was not highlighted enough and could be overlooked. For example, in some cases, assessors would tick yes for domestic violence but would not provide the supporting evidence.[27] It was also suggested that more in-depth questions could focus on issues relating to stalking/harassment. One interviewee said that OASys focused upon partner violence and was less tailored towards violence that takes place in a domestic setting between other family members, for example, mother and son or other non-intimate partner.[28]

[25] RM2000 is a risk measurement tool specifically designed to assess risk for male sex offenders (Thornton, 2007).

[26] Spousal Assault Risk Assessment (SARA) is a clinical checklist of risk factors for spousal assault. It is used in the UK prison and probation service to determine whether an offender is suitable for an intervention (Kropp, Hart, Webster and Eaves, 1995).

[27] The OASys assessment provides free text boxes to allow the assessor room to expand on any issues.

[28] The current question in the OASys assessment asks for evidence of domestic violence which relates to any form of violence, or emotional or physical abuse, threatened or actual, that occurs between two domestic partners. The question is deliberately narrow to identify a specific group of offenders and does not include physical violence against other relatives, or members of the household.

Positive/protective factors

Around three-fifths (61%) of assessors felt that no positive or protective factors were missing from the OASys layers (Table 2.4). Some of those who disagreed (16% of the sample) felt that OASys assessments largely focused on offenders' problems and their risk factors rather than their strengths (see also Chapter 6*).* The following positive/protective factors were mentioned in assessors' open-ended responses:

- Accommodation
- Family support / stable relationships
- Employment
- Financial stability
- Faith / Religion
- Cultural factors
- Educational ability / courses completed
- Self efficacy / self belief / self esteem
- Motivational factors (previous engagement with partner agencies, potential engagement with changing process)
- Offender's hobbies / interests

Diversity issues

Approximately half of the assessors felt that OASys assessments covered age (49%), gender (49%), ethnicity (49%), disability (52%) and religion (46%) very or fairly well. For sexual orientation, the proportion was nearer one third (36%; Table 2.4). There were some significant differences in the views of prison and probation assessors, with the former more likely to agree that gender, ethnicity, religion and sexual orientation were very or fairly well covered compared to their probation counterparts.

OASys training and guidance

Training

At the time of the research, training requirements varied depending on geographical location and between probation and prison, although the expectation on the probation side was that assessors should attend OASys refresher training every three years. As shown by Table 2.5, approximately a half (51%) of all respondents to the questionnaire last received OASys training two or more years ago. The remainder had received training in the last 12 months (25%) or between a year and two years ago (24%). There were significant differences between prison and probation assessors, with more prison assessors than probation assessors having completed OASys training two or more years ago (62% and 49% respectively). In terms of the availability of training, over a half (53%) of prison assessors felt that OASys training was not available when it was needed, compared with just over a third (36%) of probation assessors.[29] Importantly, current training arrangements are different between the two services. For the prison service, training is managed centrally; whereas for the probation service, training is managed by the individual trusts.

Table 2.5: Assessors' views on OASys training

Question	**Unweighted n**	**Answer (%)**		
		Less than 12 months ago	**12 months to less than 2 years ago**	**2 or more years ago**
When was the last time you received OASys training?	921	25	24	51
		Yes	**No**	-
Do you feel that OASys training is available when it is needed?	1,093	62	38	-
		Very/fairly well	**Neither well or not well**	**Not very/not at all well**
How well do you feel that OASys training meets your needs as an OASys assessor?	1,093	61	19	20

Such differences were also evident through the OASys lead interviews. In some probation trusts, the interviewees stated that most types of training (including the initial induction training and quality assurance role training) could be accessed quickly; assessors did not have to wait for a course to be fully subscribed and they would receive training within a week of starting their roles. In contrast, prison interviewees reported variability in: (i) the availability of the different types of training; and (ii) the waiting times.[30] One OASys lead mentioned difficulties around releasing prison staff to complete training courses, especially if the training took place off site.

"As soon as a new member of staff starts… they go on training as soon as is possible."
Probation, OASys training officer

"Well when people arrive they take ages to go on the course. You can't do OASys until you've done the training, you can't log on to the system. You can't be an assessor unless you've done the training." Prison, OASys quality assurer

Guidance

As shown by Table 2.6, over half of the OASys assessors strongly agreed or agreed that the OASys online help is sufficiently clear (56%) and sufficiently detailed (51%). Just under half (45%) of the assessors thought that there was scope for improving the OASys online help. Suggestions for

29 This survey was undertaken at a time where there was a gap in prison training due to the imminent release of the new IT system through the OASys-R project.
30 As noted above, this could be due to the gap in availability of prison OASys training at the time the survey was undertaken.

improvement included: (i) making the guide more user-friendly through improved navigation; and (ii) improvements and updates to the Appendix offence code list to make it quicker and easier to identify the relevant offences.

Table 2.6: Assessors' views on OASys guidance

Question	Unweighted n	Answer (%)		
Do you agree or disagree with the following statements?		**Strongly agree/agree**	**Neither agree nor disagree**	**Strongly disagree/disagree**
• The OASys online help guidance is sufficiently clear	1,093	56	24	20
• The OASys online help guidance is sufficiently detailed	1,093	51	27	22
• Supplementary OASys guidance issued by NOMS is sufficiently clear	1,093	37	44	19
• Supplementary OASys guidance issued by NOMS is sufficiently detailed	1,093	36	45	19
• Locally devised OASys guidance is sufficiently clear	1,093	44	38	18
• Locally devised OASys guidance is sufficiently detailed	1,093	43	33	23

There were differences in the way assessors made use of the different OASys help manuals and guidance. For example, some who did not know of the existence of the online help guidance or had never used it, chose instead to refer to the OASys manual or ask colleagues for support and clarification when needed.

> *"I have been with my area's service for numerous years and nobody knows about the online help".* Probation Service Officer (Questionnaire response)

> *"I have never used the online help, and didn't know it existed. If I need help I will refer to the manual."* Probation Officer (Questionnaire response)

The interviewees similarly stated that assessors did not always use the guidance if they needed help, instead relying on locally produced guidance or preferring to discuss the issue with their colleagues.

> *"We have additional guidance – discussion and training database – which supplements the main guidance [this includes templates with example assessments]. "* Probation, Training Officer

"That's really difficult for me to comment on because I don't use it [the online guidance]. And I'm not conscious of officers using it routinely. I think what they [assessors] do if they come unstuck is they come and talk about the issue, discuss it with fellow workers, rather than necessarily use the [online help]... I'm not conscious of them using it a lot."

Prison, Quality Assurance Lead

The majority of the OASys leads also agreed that there was scope for improving the OASys online help. Whilst some of the OASys leads/managers felt the guidance was useful, others felt it did not always provide the necessary answers. Furthermore some felt it was cumbersome and difficult to navigate through.[31]

OASys quality assurance procedures

As set out in Table 2.7, approximately two thirds of assessors (65%) felt that the level of quality assurance for OASys assessments was about right. There were significant differences between probation grades, with over three quarters of PSOs (79%) holding this view, compared to two thirds of POs (66%).

Table 2.7: Assessors' views on the OASys quality assurance procedures

Question	**Unweighted n**	**Answer (%)**		
		Too little	**About right**	**Too much**
Overall, I feel that the level of Quality Assurance (QA) in OASys assessments is:	1,093	18	65	17
		Very / fairly helpful	**Neither helpful or unhelpful**	**Not very / not at all helpful**
How helpful do you think the introduction of the QA process[32] has been at raising the quality of assessments?	1,093	56	22	22
		Excellent / very good	**Good**	**Fair / Poor**
Overall, how would you personally rate:				
the quality of OASys assessments in YOUR trust/establishment?	1,093	44	34	22
the quality of OASys assessments in OTHER trusts/establishments?	1,093	29	31	40
		Yes	**No**	
Do you believe there is scope for improving the quality of OASys assessments?	1,093	75	25	-

[31] The navigation of the guidance has been improved through the OASys-R project. Once within an assessment, the application now has a navigation menu, which moves with the user as they scroll up or down the page.
[32] This relates to the Quality Assurance role which was introduced in 2010.

Over half (56%) of the assessors thought that the introduction of the quality assurance process had raised the quality of OASys assessments. Over two-fifths (44%) of assessors rated the quality of OASys assessments in their own trusts/establishments as excellent or very good (Table 2.7). Prison assessors were more likely to hold this view compared to their probation counterparts (57% and 42% respectively). Approximately a third of assessors (29%) rated the quality of OASys assessments in other trusts/establishments as excellent or very good (Table 2.7). Within the interviews, the majority of OASys leads agreed that assessors were often more critical of assessments completed at other trusts/establishments. They attributed it to 'human nature' and assessors' critical attitudes.

Three quarters (75%) of the assessors believed that there was scope for improving the quality of OASys assessments (Table 2.7), with POs more likely to hold this view than PSOs (79% and 66% respectively). When they were asked how the quality of OASys assessments could be improved, frequently mentioned themes included:

- Management and supervision
    - Increased support from managers and more constructive feedback.
    - Building in a culture of developmental supervision on a monthly basis.
    - Improved accountability for locking and signing off assessments which are blank or not fully completed.
    - Encouraging assessors to use their assessment rather than their 'story telling' skills – some assessors tended to provide a narrative or description of offenders' lives rather than an analysis of their risk/needs factors.
    - Reducing caseloads/targets.
- Training and guidance
    - Updating training packages to reflect all recent changes to OASys assessments.[33]
    - Reviewing whether staff needed to attend group or one-to-one training.
    - Having more, better, consistent and regular training across both services.
- Revisions to OASys
    - Further streamlining of the assessment.
    - Improving the RoSH content, e.g. a clearer distinction between risk of harm and risk of serious harm.
- Reviewing QA processes
    - Developing a feedback questionnaire asking assessors how beneficial the QA process has been.
    - Conducting benchmarking exercises on cases within the QA database, getting others to quality assure the same assessments to check the consistency of ratings.[34]

[33] This has been addressed through the OASys-R project.
[34] The quality assurance of others' assessments already takes place in many regions.

2.4 Implications

The findings presented in this chapter demonstrate that OASys was seen by assessors as having various strengths. Crucially, the majority of assessors felt that the information recorded in an OASys assessment supported them well in managing offenders' risks and needs and enabled them to monitor progress and change over time. Feedback on the RoSH ratings was positive, providing support to their use in the new Case Allocation System (CAS) for allocating cases to the National Probation System (NPS) or Community Rehabilitation Companies (CRCs). However, the assessors' responses also indicated areas for potential improvement. More specifically, consideration should be given to the following:

The assessment process:

- Improving assessor awareness of: (i) the value and workings of the actuarial reoffending predictors; and (ii) the current targeting criteria for the OASys layers and fast reviews. The former will be important when designing the communications and training documents which accompany the implementation of the next iterations of the reoffending predictors (see Chapters 8 and 9).
- Encouraging establishments and trusts to share information/good practice, e.g. how to access offender information.

OASys targeting:

- Revisiting the targeting of OASys and its layers. As the resources available for assessment appeared to be stretched, any recommendations regarding future targeting will need to pay careful regard to the resource implications.

OASys content:

- Removing areas of duplication within OASys (e.g. between the RoSH sections), and revisiting the structure and content of the OASys self assessment questionnaire and the sentence plan.
- Further examining the inclusion of positive and protective factors (see also Chapter 6).

OASys guidance and training

- Improving awareness of the online help guidance alongside the OASys manual.

OASys quality assurance processes

- Encouraging the set up of local initiatives, such as staff briefings, discussion forums, peer review support mechanisms and improved developmental supervision to help continue to improve the quality of OASys assessments.

A number of the issues which have been highlighted in the research have now been addressed through the OASys replacement (OASys-R) project and the introduction of a new OASys IT system.

For example, the navigation of the OASys online help and the positioning of the offender SAQ form. Similarly, the layer of assessment required for different offenders is currently under review as part of the Offender Management Change Programme. It is also important to recognise that the national figures disguise significant differences between the individual probation trusts and prison establishments. These differences illustrate clear potential for the alleviation of trust/establishment specific issues through improved guidance and the sharing of good practice.

3. The prediction of reoffending by age, gender and ethnicity

This chapter sets out the findings from research which tested the ability of OGP1 and OVP1 to predict proven reoffending for offenders of different gender, ethnicity and age. Key points are as follows:

- Among all offenders, actual (proven) reoffending was significantly below the predicted rate, especially for non-violent offending, reflecting known overall reductions in reoffending since OGP1 and OVP1 were created.
- Among women, non-violent reoffending was 3.7% below predicted, compared with 2.1% below predicted for male offenders. While the non-violent reoffending of White offenders was 2.6% below predicted, for Asian and Black offenders it was 3.4% and 2.2% respectively above predicted. Actual and predicted non-violent reoffending were identical for offenders aged 18–19, but actual rates were between 1% and 4% lower than predicted for all other age groups.
- Actual/predicted differences by gender and ethnicity were far smaller for violent reoffending, while violent reoffending rates were 2% above predicted at age 18–19 and at least 4% below predicted for 22–23, 46–50 and 51+ year olds.
- Both predictors achieved reasonable relative predictive validity – successfully distinguishing likely reoffenders and likely non-reoffenders – for all offender groups. Relative predictive validity was greater for female than male offenders, for White offenders than offenders of Asian, Black and Mixed ethnicity, and for older than younger offenders. After controlling for differences in risk profiles, lower validity for all Black, Asian and Minority Ethnic (BME) groups (non-violent reoffending) and Black and Mixed ethnicity offenders (violent reoffending) was the greatest concern.

Statistical modelling suggests that some between-group differences in risk factors for reoffending may exist, but incorporating these factors does not improve prediction of reoffending sufficiently to justify the introduction of separate predictors. Revision of the predictors will be required to ensure that they reflect contemporary patterns of reoffending.

3.1 Context

Awareness of diversity issues is important to NOMS's offender assessment and management practice, and they were thoroughly considered in developing the Offender Management Model through an equality impact assessment. It is therefore important to establish whether NOMS's risk predictors are equally valid for offenders with different personal characteristics. Data available in OASys allows identification of age, gender and self-reported ethnicity. The analysis in this chapter therefore studies these aspects of diversity in risk prediction.[35]

The aims of the analysis were to:

1. Establish whether the predictive scores within OASys predict absolute and relative risks of reoffending equally well for offenders of different age, gender and ethnicity.
2. Where any differences in predictive validity exist, investigate the reasons for these differences in order to:
   a. provide guidance to operational staff on use of the scores as they are now; and
   b. inform future research to update the predictive scores.

3.2 Approach

Sample

Lists of offenders assessed using OASys by 31 March 2007 were submitted to the Ministry of Justice's (MOJ's) Police National Computer (PNC) research database in June 2009. The following cases were filtered out:

- those whose index offence could not be identified on the PNC;
- those whose assessment was not within three months of their community sentence or discharge from custody;
- those for whom OGRS, OGP1 or OVP1 scores could not be calculated;[36]
- those whose follow-up commenced less than 36 months prior to the PNC extract date; and
- those included in the original OGP1/OVP1 construction and validation study (Howard, 2009).

A sample of 92,514 cases remained for the analysis of 24-month proven reoffending outcomes. These offenders commenced community sentences or were discharged from custody between July 2004 and June 2006. Offenders could be included more than once, when these assessments were related to separate non-concurrent sentences. The eligible sample comprised 24% on licence from a custodial sentence, 31% on Criminal Justice Act 2003 (CJA 2003) Community Orders, 6% on Suspended Sentence Orders and 39% on pre-CJA 2003 community sentences. The breakdown by principal current offence was as follows:

- 25% violence against the person
- 2% robbery

[35] Research on the associations between risk prediction and sexual orientation, disability and other aspects of diversity is not currently possible. While information on religious faith is also collected in OASys, levels of data completion, the relatively small size of many faith groups within the offender population and especially the strong correlation between faith and ethnicity make meaningful analysis of the relationship between faith and reoffending difficult.

[36] Due to missing date of birth or apparent convictions aged under 10, or missing data on OGP1 or OVP1 items.

- 7% public order offences
- 3% sexual offences
- 5% burglary
- 14% theft and handling
- 3% fraud and forgery
- 8% absconding
- 19% motoring offences
- 4% criminal damage
- 7% drugs offences
- 3% all other offences

Demographic details of the eligible sample are included in the results section below.

Procedure

The analysis focused on 24-month proven reoffending outcomes. The 24-month period related to the time following community sentence or discharge from custody within which reoffending must have occurred to be included in the outcome measure. An additional 12-month 'buffer' period was allowed for the offence to be brought to justice and PNC data entry to occur, summing to the 36-month period specified in the Sample section above. For convenience, the outcomes predicted by OGP1 and OVP1 are referred to as non-violent and violent proven reoffending respectively, and the term reoffending is sometimes used as a synonym for proven reoffending. The potential complexity of the relationship between true reoffending and proven reoffending should always be noted, and is discussed further in the Implications section below.

The validity of OGP1 and OVP1 across diverse groups was tested in two respects: their ability to identify absolute and relative risk. An additional benefit of the analysis presented in this chapter is that testing absolute and relative predictive validity on this sample also provides information on how well OGP1 and OVP1 predict for offenders post-dating the sample on which they were constructed and validated. As Howard (2009) reports, OGP1 and OVP1 were developed on samples assessed until September 2004; this chapter's eligible sample covers the period October 2004 – June 2006.

To test absolute risk, actual and predicted proven reoffending rates are compared for all offenders, and for those in given score bands. The bands initially covered 5-point ranges along the 100-point OGP1 and OVP1 scales, with adjacent bands merged where necessary to increase numbers. The figures which present these results include only bands containing at least 50 offenders, in order to avoid presenting potentially misleading results based on very small numbers.

To test relative risk, Area Under Curve (AUC) statistics are presented. High AUCs are clearly desirable if a risk predictor is to help offender managers and other staff correctly identify the subset of offenders who require scarce supervision and intervention/programme resources and/or should be incapacitated (imprisoned or subjected to restriction on their movements or activities while in the community) for public protection. It is important to understand that high AUCs arise when many offenders have very low and/or very high probabilities of reoffending and the risk predictor accurately estimates these probabilities. Lower AUCs arise when groups are homogeneous (i.e. offenders within the group have similar actual probabilities to one another) or the actual probabilities of reoffending are less extreme (i.e. closer to 50%) than predicted because the predictor fails to include relevant risk factors and/or under- or overestimates the importance of the risk factors it does include.

Differences between AUCs for the same predictor are detected using T-tests (Gönen, 2007), with reference groups of age 26–30, male and White offenders (these are the most frequent age, gender and ethnic groups). If AUCs differ significantly between groups, this may be because the predictor fails to discriminate effectively between higher and lower risk offenders, or it may be that the group is unusually homogenous. AUCs will be lower when offenders do not differ on major risk factors. Accordingly, AUCs for specific age and gender groups are likely to be lower than the overall AUCs, as age and gender are each scored in both OGP1 and OVP1. The banded score figures mentioned above are useful for showing the extent of differences in proven reoffending across bands between those in each group. A more precise method is the calculation of standardised AUCs. Differences in the proportions with each score are controlled for by applying each score's reoffending rate among the group of interest to the population distribution of the reference group. This means that only the reoffending rates are varied between the group of interest and the reference group.[37]

Note that age is categorised using the scheme applied in the original OGP1/OVP1 study (Howard, 2009), and ethnicity uses broad rather than Census groups. Both of these categorisation schemes have the practical benefit that most sub-groups are large enough that the confidence intervals around AUC estimates are reasonably narrow and each sub-group can itself be broken down into smaller groups in order to study reoffending rates by levels of OGP1 and OVP1 score. However, the 'other' ethnicity group does remain small (n=503), and results for this group therefore will be presented in tables and figures but rarely referred to in the text.

Where appropriate, logistic regression models were run to predict non-violent or violent reoffending among a specific group of offenders. This helps to identify whether this group differs from the overall patterns in terms of the associations between static or dynamic risk factors and reoffending, and whether there is sufficient value to be gained from the development of differing scoring systems.

[37] Where the group of interest includes scores which are not present in the reference group, outlying scores are combined (e.g. no male offenders score below 5 on OVP1, so the combined reoffending rate of female offenders scoring 0-5 is applied to the population of males scoring 5 when calculating the standardised AUC for females). When the group of interest has no observations for scores which are present in the reference group, the reoffending rate of the reference group is used (e.g. no female offenders score 85 on OVP1; the male reoffending rate for those scoring 85 is therefore used in the female standardised AUC calculation).

3.3 Results

Table 3.1 presents the predictive validity of OGP1 and OVP1 as estimates of absolute risk among all offenders and each sub-group. Among all offenders, actual reoffending was significantly below the predicted rate, especially for non-violent offending. This reflects the gradual reduction in proven reoffending across the entire NOMS caseload during the past decade (Ministry of Justice, 2013a). Among women, non-violent reoffending was a further 1.6 percentage points lower than predicted, compared with male offenders (i.e. the actual rate for women was 3.7 percentage points below predicted, whereas for men it was 2.1 percentage points below predicted). White offenders reoffended 2.6 percentage points below predicted, whereas Asian and Black offenders reoffended 3.4 and 2.2 percentage points above predicted respectively. Actual/predicted differences by gender and ethnicity were far smaller for violent reoffending.

Actual and predicted non-violent reoffending were identical for offenders aged 18 or 19, but actual rates were between 1.3 and 4.0 percentage points lower than predicted for all other age groups. Violent reoffending rates were 1.7 percentage points above predicted for 18–19 year olds, whereas they were at least four percentage points below predicted for 22–23, 46–50 and 51+ year olds. The overall age pattern for violent reoffending shows a more extreme age trend than the OVP1 scoring algorithm allows for. Howard (2009) explains that the OVP1 scoring system was deliberately simplified in order to make the weighting system more user-friendly and increase the scope for change on dynamic factors to influence reoffending. However, this was shown to have very little effect on the overall predictive validity of OVP1 and had a very mild impact on age, changing its weight from 23 to 20 of the 100 points, which cannot account for the largest residuals here, and especially not for the large negative residual for the relatively young 22–23-year-old age group. (That is, the artificially lowered age weighting should, if anything, cause the predicted violent reoffending rate of 22–23 year olds to be too low rather than too high.) Overall, the age trends here suggest that reoffending rates fell between 2002–4 and 2004–6 among all age groups except those aged 18–19 years old.

Table 3.1: Absolute predictive validity: Actual and predicted non-violent and violent reoffending by gender, ethnicity and age

Offender group *(number / % of offenders)*	Non-violent reoffending within 2 years of sentence/discharge			Violent reoffending within 2 years of sentence discharge		
	Actual	**Predicted**	**Residual (actual minus predicted)**	**Actual**	**Predicted**	**Residual (actual minus predicted)**
All offenders (92,514)	***40.4%***	***42.7%***	***-2.3%***	***28.7%***	***29.8%***	***-1.1%***
Gender						
Female *(12,194 / 13.1%)*	35.8%	39.5%	-3.7%	19.5%	19.8%	-0.3%
Male *(80,320 / 86.9%)*	41.0%	43.1%	-2.1%	30.1%	31.3%	-1.2%
Ethnicity						
Asian *(2,661 / 2.9%)*	36.2%	32.8%	3.4%	21.5%	22.0%	-0.5%
Black *(3,361 / 3.6%)*	42.8%	40.6%	2.2%	27.0%	26.7%	0.3%
Mixed *(1,485 / 1.6%)*	47.6%	47.9%	-0.3%	32.1%	32.5%	-0.4%
Other *(503 / 0.5%)*	31.2%	34.4%	-3.2%	17.3%	20.4%	-3.1%
White *(75,006 / 81.1%)*	41.5%	44.1%	-2.6%	29.5%	30.6%	-1.1%
Missing / not stated *(9,498 / 10.3%)*	30.7%	34.4%	-3.7%	25.2%	26.5%	-1.3%
Age						
18–19 *(10,379 / 11.2%)*	51.6%	51.5%	0.1%	45.5%	43.8%	1.7%
20–21 *(10,662 / 11.5%)*	46.7%	50.1%	-3.4%	38.5%	39.1%	-0.6%
22–23 *(8,819 / 9.5%)*	45.8%	47.4%	-1.6%	34.3%	38.7%	-4.4%
24–25 *(8,233 / 8.9%)*	45.8%	48.4%	-2.6%	30.2%	30.9%	-0.7%
26–30 *(15,625 / 16.9%)*	45.6%	47.7%	-2.1%	28.0%	28.0%	0.0%
31–35 *(13,412 / 14.5%)*	40.6%	43.1%	-2.5%	25.0%	26.5%	-1.5%
36–40 *(10,582 / 11.4%)*	32.8%	36.2%	-3.4%	22.3%	24.9%	-2.6%
41–45 *(6,896 / 7.5%)*	26.6%	29.6%	-3.0%	18.3%	21.8%	-3.5%
46–50 *(3,710 / 4.0%)*	20.0%	24.0%	-4.0%	14.2%	18.2%	-4.0%
51+ *(4,196 / 4.5%)*	13.5%	14.8%	-1.3%	8.7%	13.3%	-4.6%

Table 3.2 examines relative predictive validity, showing the AUC statistics within each offender group. Overall, the AUCs for all offenders are only fractionally below those reported for the initial (2002–04) validation sample in Howard (2009) – a positive finding, given the increasing homogeneity of assessed offenders as the scope of OASys gradually narrowed due to the withdrawal of assessments from offenders with Unpaid Work-only requirements in some probation areas.

As in Howard (2009), both tools had significantly higher AUCs for female than male offenders. AUCs for Asian, Black and Mixed ethnic groups were significantly lower than those for White offenders for both tools, though the difference between those of Asian and White ethnicity was only marginally significant for OVP1. The relative validity of both tools was greater at older ages. The AUCs of OGP1 were lower among those aged under 26 than those aged 26+, with similar AUCs for all age groups above this threshold and the very lowest AUCs found among the very youngest offenders. The AUCs of OVP1 were quite uniform among offenders aged 18–35 and greater for older offenders, with an especially high score for those aged 51+.

Table 3.2: Relative predictive validity: Area Under Curve (AUC) statistics by gender, ethnicity and age

Offender group *(number of offenders)*	**OGP1** AUC			**OGP1** T-test (vs. reference group 26–30/Male/White)		**OVP1** AUC			**OVP1** T-test (vs. reference group 26–30/Male/White)	
	Estimate	***Lower CI***	***Upper CI***	***T***	***p value***	***Estimate***	***Lower CI***	***Upper CI***	***T***	***p value***
All offenders (92,514)	***0.794***	***0.791***	***0.797***			***0.743***	***0.740***	***0.747***		
Gender										
Female *(12,194)*	0.822	0.815	0.830	63.602	0.000	0.763	0.752	0.773	82.081	0.000
Male *(80,320)*	0.789	0.786	0.792	-	-	0.737	0.733	0.740	-	-
Ethnicity										
Asian *(2,661)*	0.767	0.749	0.785	10.136	0.001	0.723	0.701	0.746	3.079	0.079
Black *(3,361)*	0.741	0.724	0.757	42.596	0.000	0.713	0.694	0.732	10.356	0.001
Mixed *(1,485)*	0.744	0.720	0.769	17.115	0.000	0.707	0.679	0.735	6.862	0.009
Other *(503)*	0.752	0.705	0.798	3.614	0.057	0.738	0.683	0.793	0.048	0.826
White *(75,006)*	0.797	0.794	0.800	-	-	0.744	0.741	0.748	-	-
Missing / not stated *(9,498)*	0.784	0.774	0.794	6.227	0.013	0.745	0.733	0.756	0.002	0.967
Age										
18–19 *(10,379)*	0.741	0.731	0.750	64.803	0.000	0.702	0.692	0.712	0.095	0.757
20–21 *(10,662)*	0.751	0.742	0.760	40.664	0.000	0.712	0.703	0.722	3.310	0.069
22–23 *(8,819)*	0.764	0.754	0.774	15.991	0.000	0.710	0.699	0.721	1.844	0.175
24–25 *(8,233)*	0.771	0.761	0.781	8.252	0.004	0.707	0.695	0.719	0.908	0.341
26–30 *(15,625)*	0.789	0.782	0.796	-	-	0.700	0.691	0.709	-	-
31–35 *(13,412)*	0.795	0.787	0.802	1.211	0.271	0.706	0.696	0.716	0.728	0.394
36–40 *(10,582)*	0.788	0.778	0.797	0.035	0.853	0.736	0.725	0.747	24.686	0.000
41–45 *(6,896)*	0.799	0.788	0.811	2.252	0.133	0.739	0.724	0.754	19.715	0.000
46–50 *(3,710)*	0.791	0.772	0.809	0.037	0.847	0.755	0.733	0.777	20.385	0.000
51+ *(4,196)*	0.780	0.759	0.800	0.687	0.407	0.806	0.784	0.827	77.560	0.000

Absolute and relative predictive validity by gender

Figure 3.1 presents actual non-violent reoffending rates by gender and OGP1 score band. The results controlling for score bands are consistent with the overall finding that non-violent reoffending by female offenders was overestimated to a greater extent than non-violent reoffending by male offenders. While overall overestimation was 1.6 percentage points greater for females than males, the equivalent differences within score bands ranged from +5 to -2 percentage points.

There was little difference between the slope of the male and female reoffending curves, suggesting that differences in the AUCs were due to differences in population distribution rather than in true relative predictive validity. Figure 3.2 confirms this, presenting the distributions of OGP1 scores. Female offenders had a peak in OGP1 scores between 8 and 16, with 16% of offenders having these scores, associated with estimated non-violent reoffending rates of just 7 to 11%. In all, 33% of female offenders had OGP1 scores of no more than 26 (i.e. estimated non-violent reoffending rates below 20%), compared with 24% of male offenders, despite their comparatively similar overall estimated rates. The standardised AUC was calculated at 0.793, just 0.004 above that for male offenders, confirming that the higher AUC of OGP1 for female offenders was almost wholly due to their score distribution.

Figure 3.3 presents actual violent reoffending rates by gender and OVP1 score band. This suggests that OVP1 may slightly overestimate female rates at most scores. This pattern reverses at higher scores (46 and above) but, as Figure 3.4 shows, few female offenders had such scores. The crossing pattern of the two curves suggests that OVP1 may predict relative risk less well for females than males once population distribution has been controlled for. Standardisation confirms this, producing an AUC of 0.720, compared with the unstandardised female AUC of 0.760 and the male AUC of 0.737. Producing a separate predictor just for women has very modest success when a naive modelling approach is used, improving the AUC to 0.771: it includes education, relationship and drug misuse items while excluding accommodation and employability. A more thorough approach – splitting female offenders between model construction and validation datasets – fails, with an AUC of 0.758 for the female-only model and 0.754 for OVP1 on the validation dataset, there being no statistically significant difference between the two (p=.11, using the method of Gönen, 2007). This suggests that the risk factors for violent reoffending among females are insufficiently different to those for males to be usefully separated in a predictive model.

In summary, the greater spreads of OGP1 and OVP1 scores among female offenders (standard deviations of 22.3 and 13.7 respectively, compared with 20.3 and 13.1 for males) means that the highest- and lowest-scoring female offenders had very well differentiated reoffending rates. However, the relative predictive validity of OVP1 – its ability to differentiate the risks of reoffending of two offenders with a given pair of scores – was slightly weaker among women than men.

Figure 3.1: Non-violent reoffending within 24 months of sentence/discharge, by grouped OPG score and gender

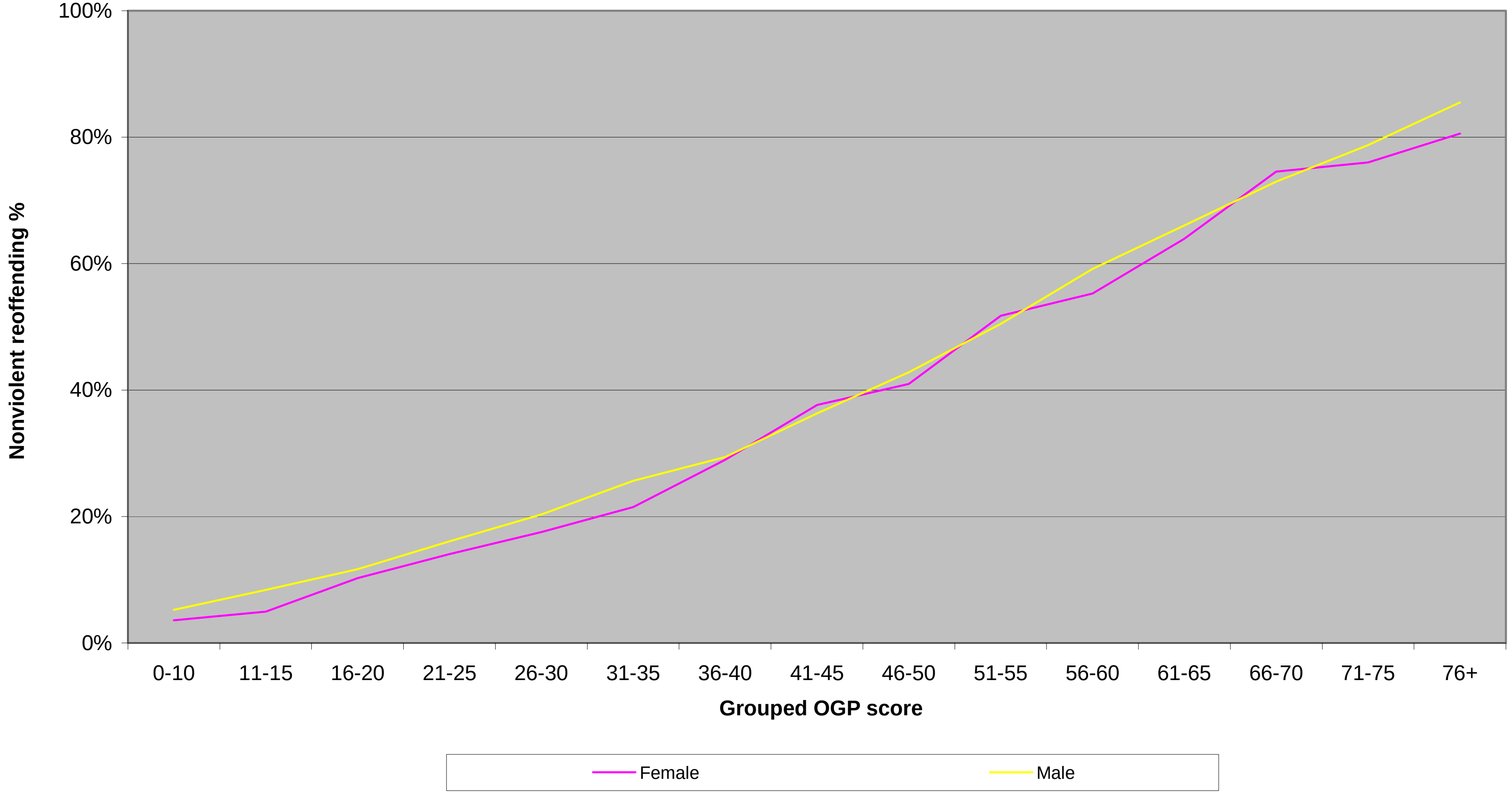


Figure 3.2: Distribution of OGP scores by gender

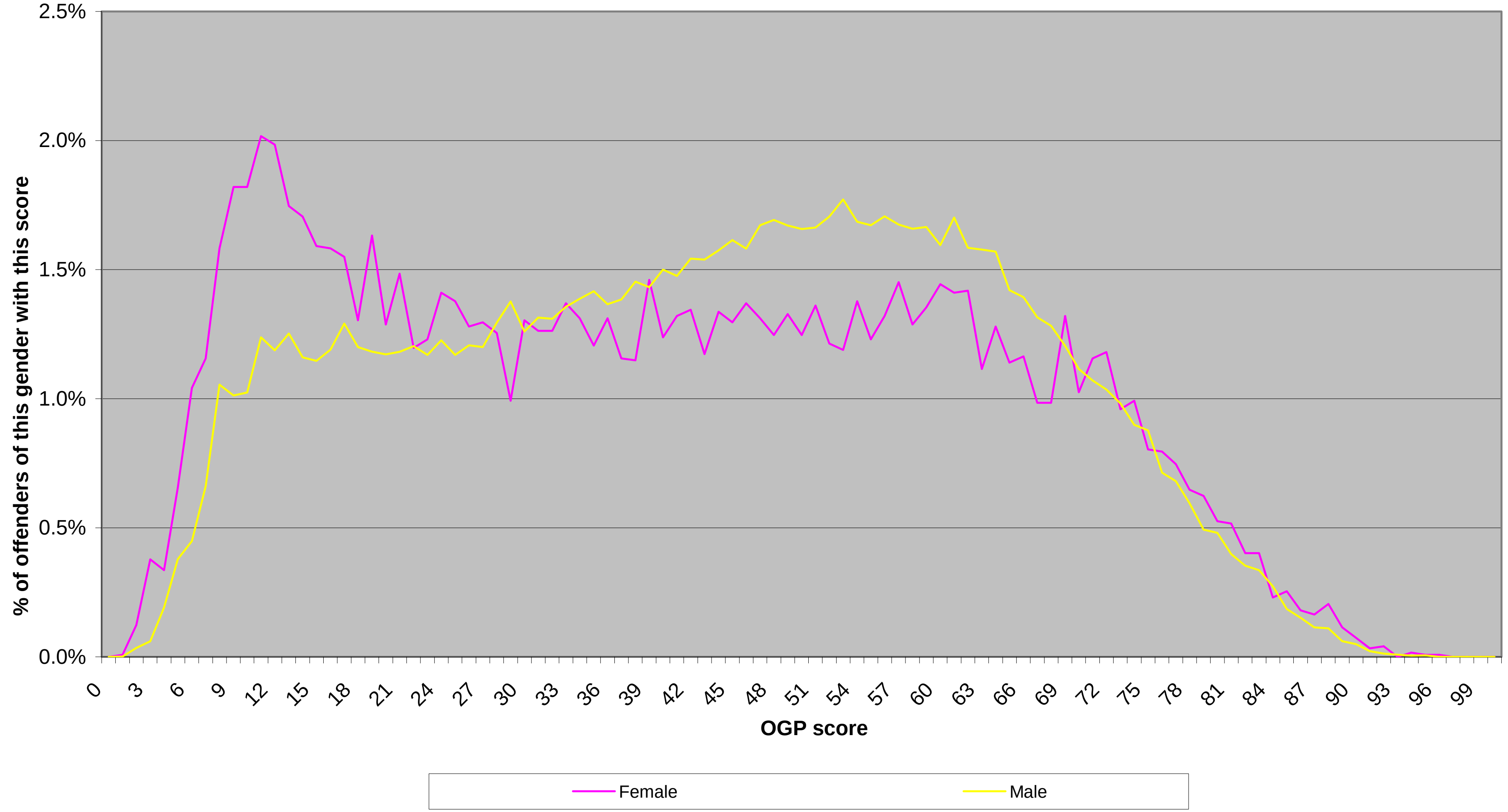


Figure 3.3: Violent reoffending within 24 months of sentence/discharge, by grouped OVP score and gender

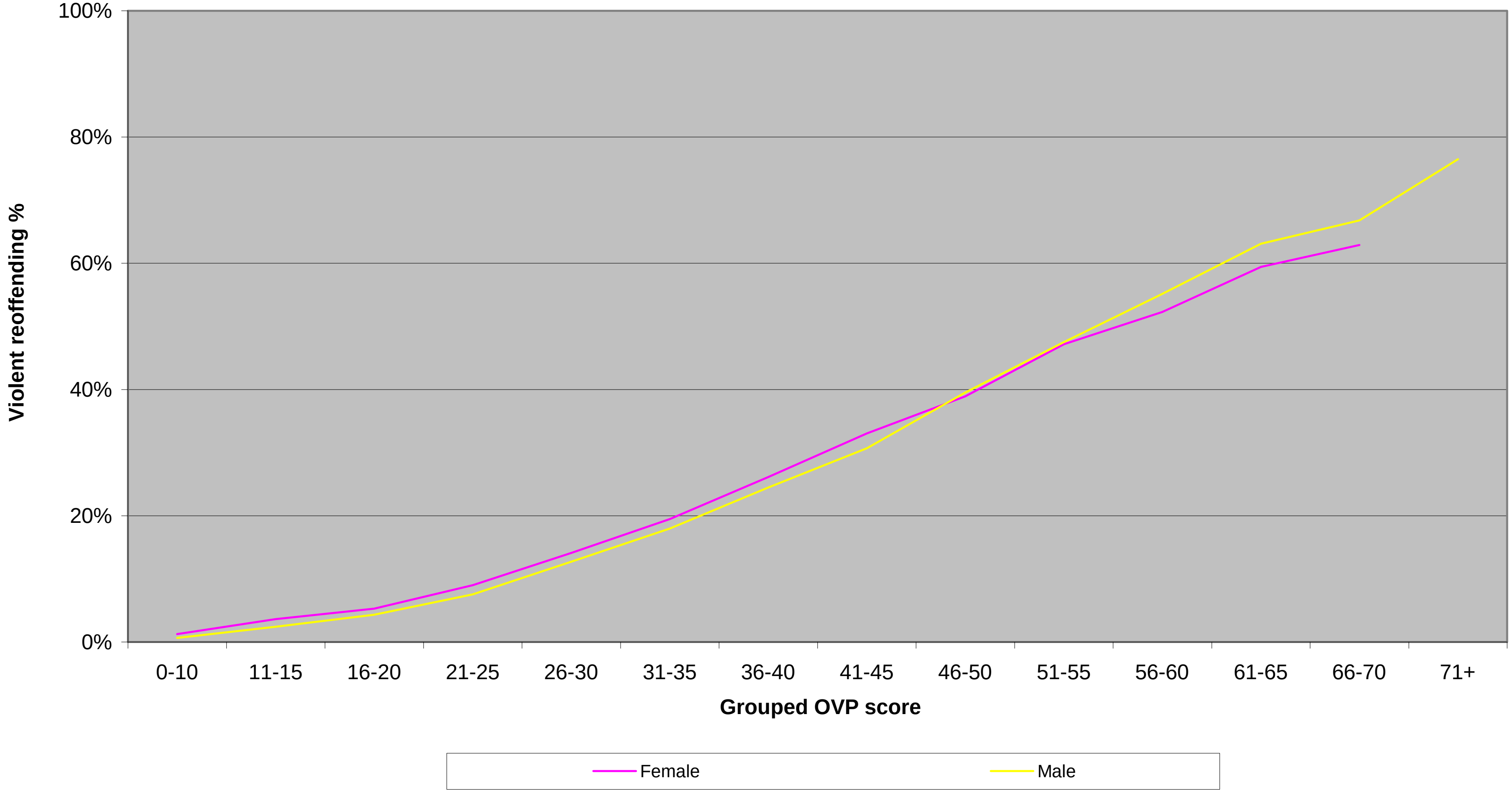
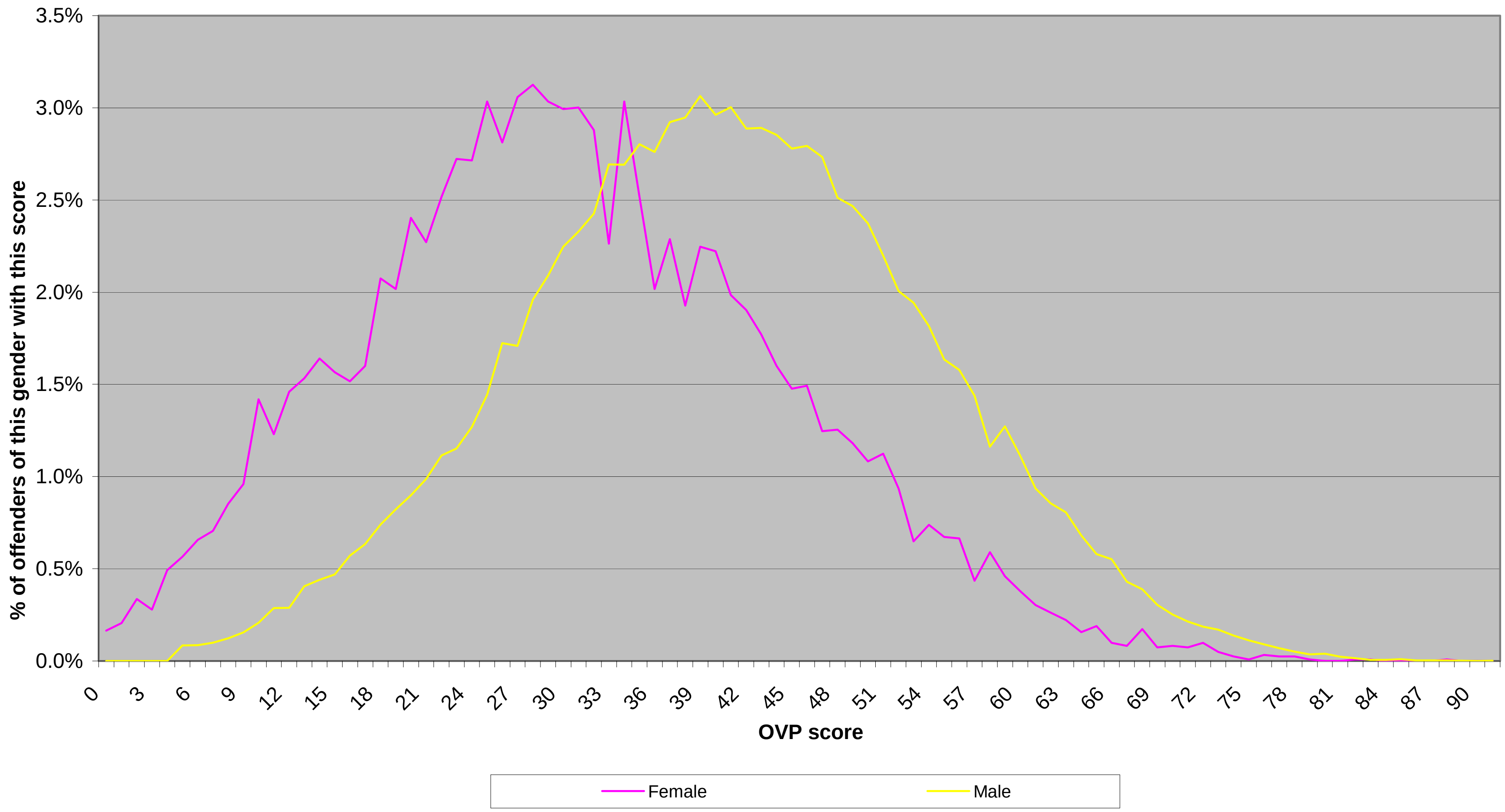


Figure 3.4: Distribution of OVP scores by gender



Absolute and relative predictive validity by ethnicity: non-violent reoffending

Figure 3.5 presents actual non-violent reoffending rates by ethnicity and OGP1 score band. The reoffending rates of Asian and other ethnicity offenders were greater than those of White offenders for most or all of the bands for which comparisons are presented, while those of Black and Mixed ethnicity offenders were greater than those of White offenders at low OGP1 scores but equal or less than those of White offenders at high OGP1 scores. The greatest differences were found at low-medium scores: Asian offenders scoring 21–25 (n=210) had a rate 10.5 percentage points above that of similar White offenders, while Black offenders scoring 26–30 (n=269) had a rate 10.1 percentage points above that of similar White offenders.

Table 3.3 presents a simple logistic regression model which confirms some of these findings. OGP1 was a significant predictor of non-violent reoffending, and Asian, Black, Mixed and other ethnicity were all associated with significant increases in the probability of reoffending. However, for Black and Mixed ethnicity offenders, the increase in reoffending probability was significantly greater at lower scores than higher scores. At an OGP1 score of 20, for example, a Black offender would have odds of reoffending 1.69 times those of a White offender, whereas at an OGP1 score of 60 the equivalent ratio would be only 1.09.[38] This means that OGP1 had less relative predictive validity within the Black and Mixed ethnicity sub-groups. Note also the difference between individual and group results: overall proven non-violent reoffending rates were lower for Asian and other ethnicity offenders than White offenders, as shown in Table 3.1, but individuals with a given OGP1 score were more likely to have a proven reoffence if they were Asian or other ethnicity.

Table 3.3: Logistic regression model predicting proven non-violent reoffending within 24 months of sentence/discharge by OGP1 score and ethnicity

Risk factor	Parameter estimate	Standard error of estimate	P value	Odds ratio (for OGP1, per point)
OGP1 score	.0623	.0005	<.001	1.064
Asian ethnicity	.451	.121	<.001	1.570
Black ethnicity	.748	.112	<.001	2.113
Mixed ethnicity	.603	.185	.001	1.828
Other ethnicity	.513	.257	.046	1.670
Ethnicity missing / not stated	.032	.073	.662	1.033
OGP1 score (Asian offenders only)	-.0027	.0029	.343	0.997
OGP1 score (Black offenders only)	-.0109	.0024	<.001	0.989
OGP1 score (Mixed ethnicity offenders only)	-.0096	.0036	.007	0.990
OGP1 score (Other ethnicity offenders only)	-.0061	.0064	.337	0.994

38 The ratio at a score of 20 = 2.113 * (0.989^{20}) = 1.69. At a score of 60, it equals 2.113 * (0.989^{60}) = 1.09.

Risk factor	Parameter estimate	Standard error of estimate	P value	Odds ratio (for OGP1, per point)
OGP1 score (ethnicity missing offenders only)	-.0024	.0016	.123	0.998
Constant	*-3.256*	*.026*	--	--

A further model disaggregated OGP1 into the OGRS3 score and the 40-point dynamic risk factor score. The weights for OGRS3 and the dynamic score were fairly similar to each other, and the simple ethnicity terms were similar to those in Table 3.3. The results suggest that there was a negative interaction with dynamic risk score (i.e. less relative predictive validity) for Asian and Mixed ethnicity offenders, while there were negative interactions with both elements of the score for Black offenders. In other words, a high score on dynamic risk factors is less predictive of proven reoffending for Asian and Mixed ethnicity offenders, and a high score generally is less predictive of proven reoffending for Black offenders, though offenders of all three groups are more likely to proven-reoffend than White offenders before taking their score into account.[39]

Further logistic regression models were run, predicting non-violent reoffending for Asian, Black and Mixed ethnicity offenders in turn. All OGP1 risk factors were included, together with other questions from the Analysis of Offences and dynamic risk factor sections which are not scored in OGP1.[40] While these models of non-violent reoffending among Black, Asian and Minority Ethnic (BME) offenders found some differences in predictive validity, using the overall reoffending model did not worsen relative predictive validity much, with none of the bespoke models for BME groups improving AUC by more than 0.01.

Figure 3.6 presents the distribution of OGP1 scores by ethnicity. The distributions for offenders of Mixed and, to a lesser extent, White ethnicity had negative skew. The distribution for Black offenders was not skewed, while those for Asian, other and missing ethnicity offenders had positive skew. Standardised AUCs were calculated for Asian, Black and Mixed ethnicity offenders, setting their score distributions to those of White offenders. The standardised AUC of OGP1 was calculated at 0.794 for Asian offenders, compared with Table 3.2's 0.797 for White offenders, but only 0.766 for Black offenders and 0.768 for offenders of Mixed ethnicity.

Together, these results suggest that OGP1 has less relative predictive validity (i.e. ability to clearly distinguish likely non-violent reoffenders from likely non-violent non-reoffenders) for offenders of Black and Mixed ethnicity than those of Asian and White ethnicity. While OGP1 also has less success in differentiating non-violent reoffenders and non-reoffenders among Asian offenders than White

39 While the geographic distribution of BME offenders was very different to that of White offenders, and reoffending rates did vary by probation area, adding probation area to the above models had no meaningful impact on the ethnicity-related results.

40 The models used forward stepwise selection at p=.05. Given the relatively small numbers in each group and the strong baseline predictive validity provided by static risk (OGRS), some risk factors with moderate associations with reoffending may not be included.

offenders, this appears to be due to the lower score distribution of Asian offenders (i.e. few Asian offenders have scores high enough to denote them as likely non-violent reoffenders) rather than a failure of OGP1 to predict well for Asian offenders of any given score. Attempts to address relative predictive validity shortcomings by creating separate models of non-violent reoffending for different ethnic groups by using different OASys dynamic risk factors were not particularly successful.

Differences in absolute predictive risk were also apparent. That is, Asian, Black, Mixed and other ethnicity offenders were all more likely to reoffend non-violently than White offenders with similar OGP1 scores. Possible explanations are discussed in the Implications section.

Figure 3.5: Non-violent reoffending within 24 months of sentence/discharge, by grouped OGP score and ethnic group

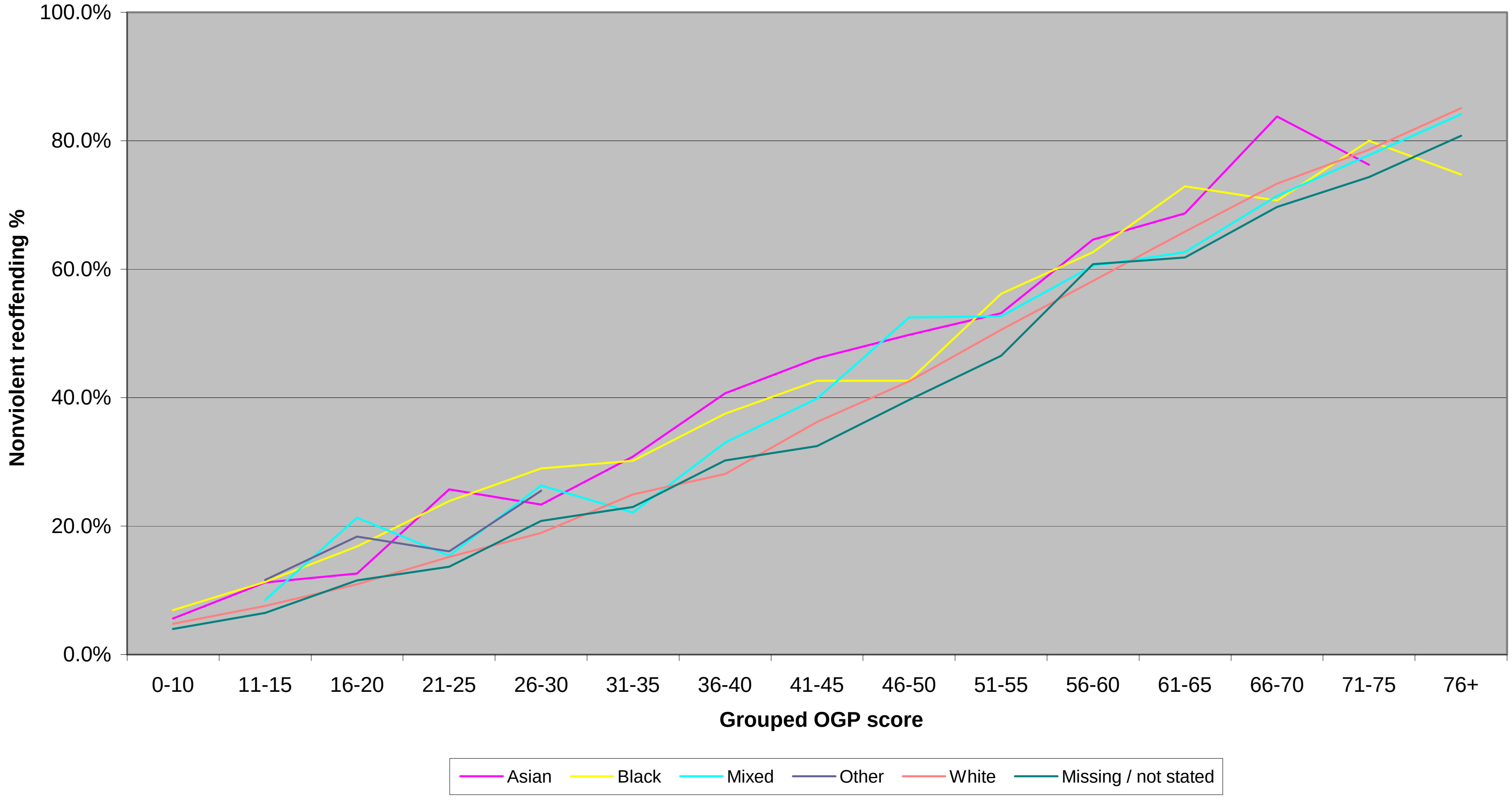
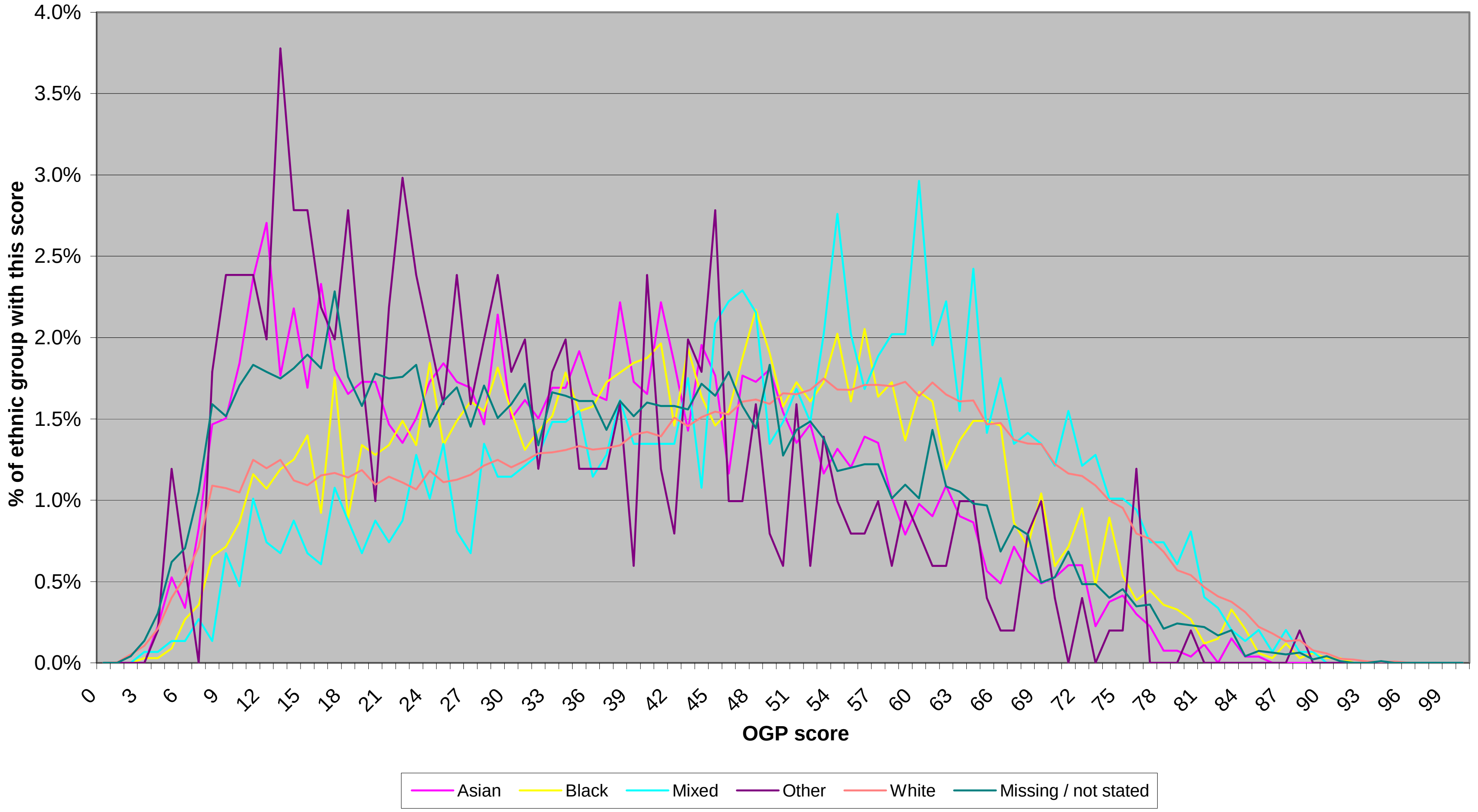


Figure 3.6: Distribution of OGP scores by ethnic group



Absolute and relative predictive validity by ethnicity: violent reoffending

Figure 3.7 presents actual violent reoffending rates by ethnicity and OVP1 score band, while Figure 3.8 shows the distribution of scores for each ethnic group. Reoffending rates for Asian, Mixed and missing ethnicity offenders fluctuated around those of White offenders, while those for Black offenders were among the highest at scores below 25 and between 36 and 50. Their scores were 5% below those of White offenders for scores of 56 to 65, but this is based on only 200 Black offenders between the two bands. Average scores were highest among Mixed and White ethnicity offenders. Few Mixed ethnicity offenders had very low scores (below 25, equating two-year probabilities no higher than 10%), whereas there were few Asian or other ethnicity offenders with scores above 50 (scores of 50 equate to two-year probabilities of 43%).

Table 3.4's logistic regression model tests whether OVP1 predicted violent reoffending with equal absolute and relative validity for offenders of different ethnicity. The patterns for Asian, Mixed, other and missing ethnicity offenders were not significantly different to those of White offenders, but the interaction between Black ethnicity and OVP1 score suggests a lack of relative validity for this group. The model suggests that the majority of Black offenders were more likely to reoffend violently than equivalent White offenders, but the 15% of Black offenders with scores of 51 and above (where the 'Black/OVP1' term exceeds the 'Black' term) were less likely to do so; this corresponds with Table 3.1's results that predicted and actual violent reoffending rates were similar for Black offenders whereas actual rates were 1% below predicted for White offenders. While the exact parameters of this model are imprecise (i.e. the 51+ crossover estimate is prone to random error), the general pattern is significant, and is similar to that for OGP1 and non-violent offending.

Figure 3.7: Violent reoffending within 24 months of sentence/discharge, by grouped OVP score and ethnic group

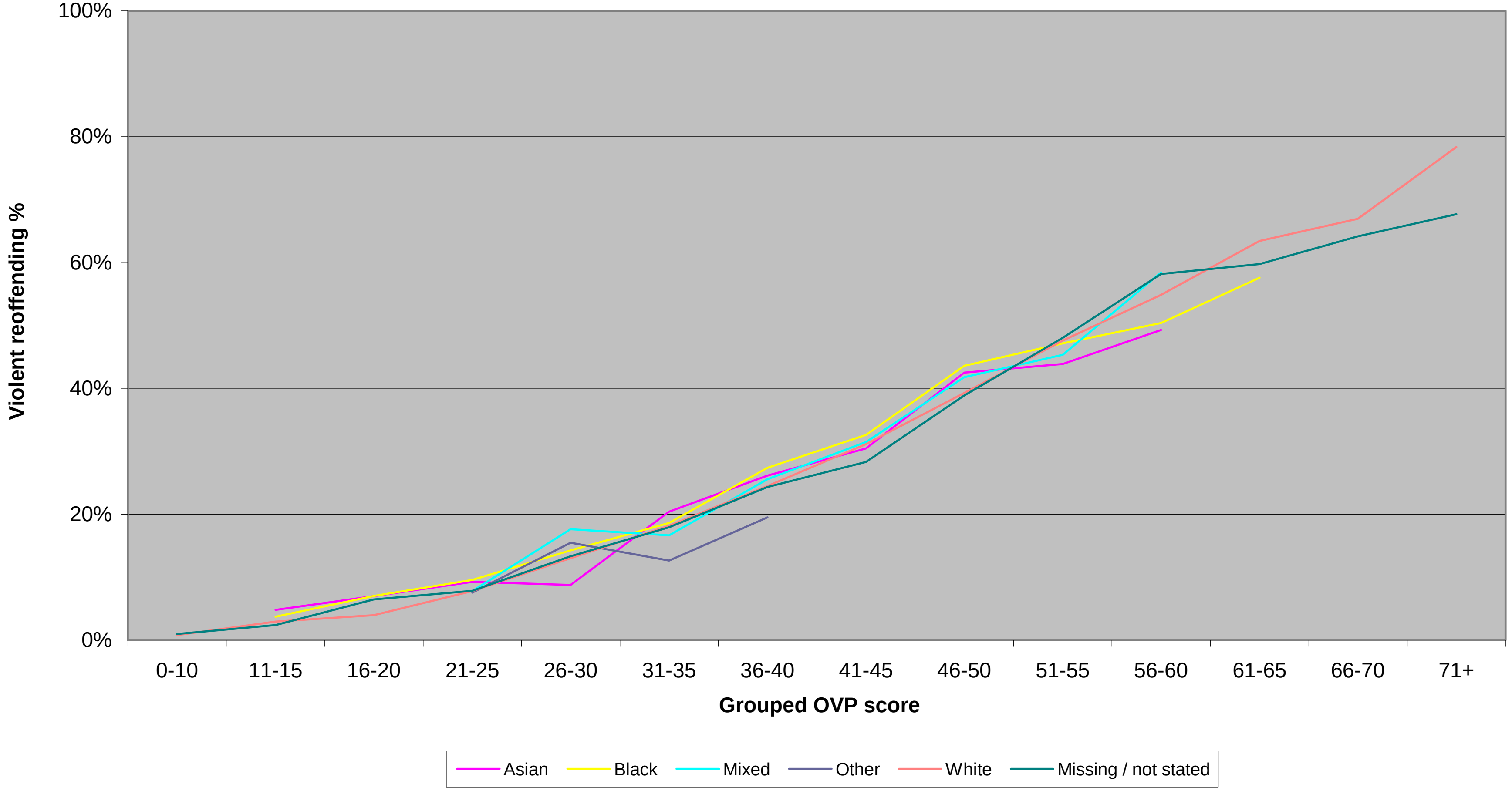


Figure 3.8: Distribution of OVP scores by ethnic group

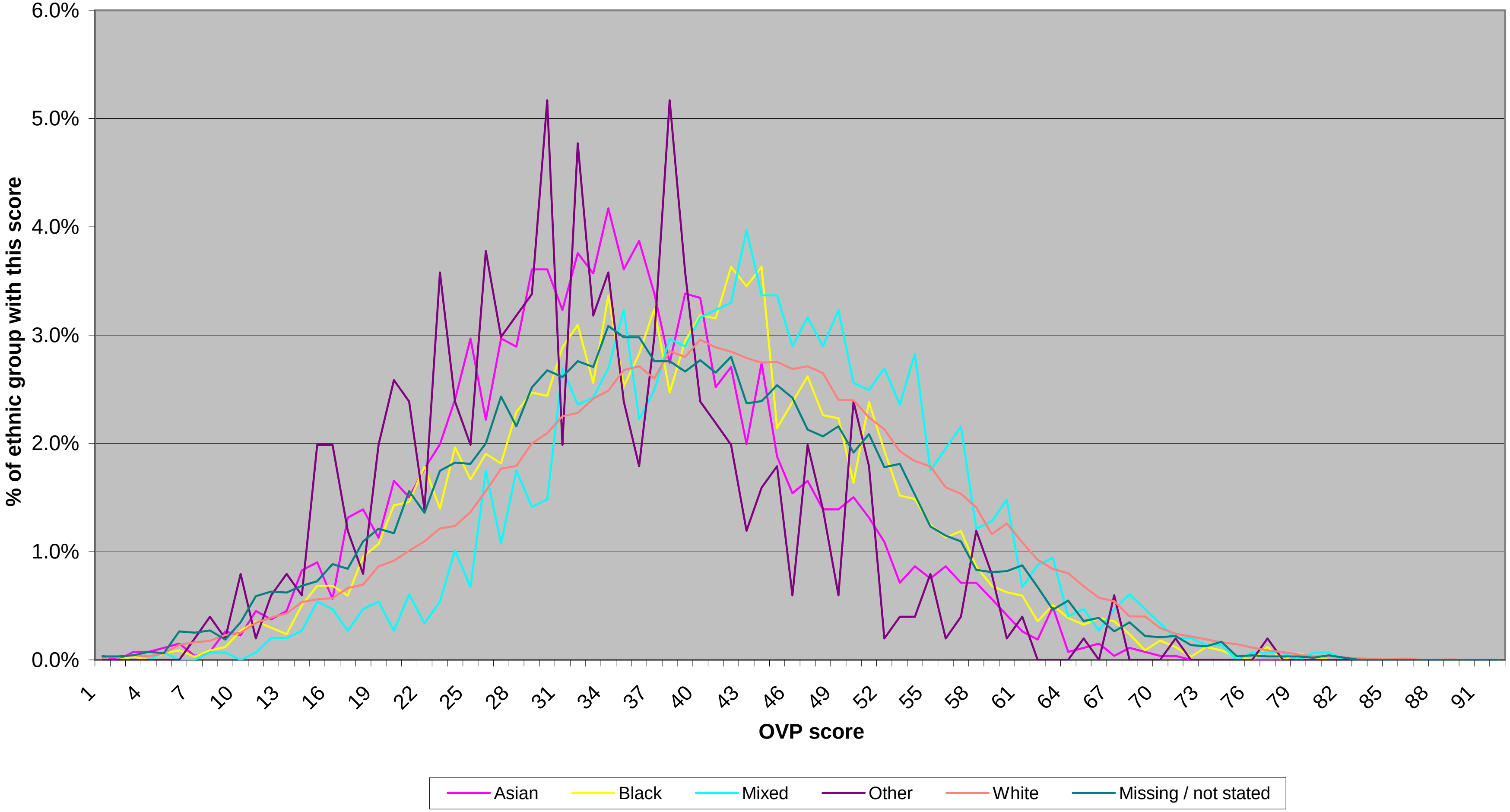


Table 3.4: Logistic regression model predicting proven violent reoffending within 24 months of sentence/discharge by OVP1 score and ethnicity

Risk factor	Parameter estimate	Standard error of estimate	P value	Odds ratio (for OVP1, per point)
OVP1 score	.0735	.00075	<.001	1.076
Asian ethnicity	-.001	.192	.997	0.999
Black ethnicity	.444	.160	.006	1.559
Mixed ethnicity	.280	.260	.280	1.323
Other ethnicity	-.288	.439	.512	0.750
Ethnicity missing / not stated	.036	.100	.722	1.036
OVP1 score (Asian offenders only)	.001	.005	.860	1.001
OVP1 score (Black offenders only)	-.009	.004	.019	0.991
OVP1 score (Mixed ethnicity offenders only)	-.005	.006	.338	0.995
OVP1 score (Other ethnicity offenders only)	.003	.011	.767	1.003
OVP1 score (ethnicity missing offenders only)	-.001	.002	.570	0.999
Constant	*-3.994*	*.034*	--	--

A further model disaggregated OVP1 for Black offenders. Static risk factors were generally more predictive than dynamic factors though this would be expected – as Howard (2009) describes, the OVP1 scoring system does artificially boost the weight of dynamic factors to encourage score changes.[41] Merely fitting this model, which reweights OVP1's existing risk factors, yields only a modest improvement, with the AUC rising from 0.713 to 0.719. Allowing a complex new model with other risk factors raises AUC to 0.727, but this is almost certainly an overestimate of what could be achieved with new offenders, and scores would be difficult to interpret. In short, refitting OVP1 for Black offenders achieves modest improvements in relative predictive validity which may not be repeatable if put into practice.

Standardised AUCs were calculated for Asian, Black and Mixed ethnicity offenders, setting their score distributions to those of White offenders. The standardised AUC of OVP1 was calculated at 0.736 for Asian offenders, 0.719 for Black offenders and 0.732 for offenders of Mixed ethnicity, compared with Table 3.2's 0.744 for White offenders. These results therefore confirm that relative predictive validity for Black offenders is the most pressing concern.

[41] Among static factors, gender and non-violent sanctions were more predictive than their current weight in OVP1 allows (i.e. males with many sanctions for non-violent offences had markedly higher reoffending rates). Among dynamic factors, failure to acknowledge the impact of offending and poor accommodation had low or negative weightings and were not significantly predictive. Attitudinal problems were by far the most predictive dynamic factor.

Absolute and relative predictive validity by age: non-violent reoffending

Figure 3.9 presents actual non-violent reoffending rates by age and OGP1 score band. Of the relatively few results which stand out visually, OGP1 appears to underestimate the reoffending likelihood of 18–19-year-old offenders at low-medium scores and, to a lesser extent, more generally. The rates of those aged 46–50 and 51+ stand out at medium scores, but few such offenders were assessed. A logistic regression model confirms that non-violent reoffending was more likely among the youngest offenders, especially at the lowest scores (i.e. being aged 18–21 significantly increased reoffending probability, as did each point of OGP1 score while the interaction between being aged 18–21 and OGP1 score significantly lowered the effect of OGP1 score on reoffending probability). The negative interaction suggests that OGP1 is less successful in sorting relative risk among younger offenders. This supports the basic message presented in Table 3.2.

Figure 3.10 illustrates the distribution of OGP1 scores by age. The oldest offenders have extremely skewed score distributions – around 55% of those aged 51+ had scores no higher than 15/100, equating to a two-year reoffending probability no higher than 10%, and few were likely to reoffend, only 5% scoring 50+/100 (50+% in 2 years). Offenders aged under 30 often scored around 50, but very few scored highly enough to be almost certain to reoffend (a score of 85+/100 is required for a 90+% two-year probability). Therefore, there is less relative difference in the likelihood of non-violent reoffending amongst the younger offenders. The very youngest offenders have the tightest central score distribution, with 42% scoring in the 41–60 range where prediction is very difficult (reoffending probabilities of 37–65%).

Removing the distributional effect by standardisation has a positive effect on prediction for the youngest offenders: standardised against the 26–30-year-old distribution, 18–19 year olds have an AUC of 0.772, considerably higher than their unstandardised AUC of 0.741 but lower than the 26–30 year olds' AUC of 0.789. The standardised AUC of 0.778 for the oldest offenders (aged 51+) is almost identical to their unstandardised AUC of 0.780.

New logistic regression models were run for the 18–19, 26–30 and 51+ age groups, to see how risk factors differed between age groups. For the 18–19 group, this bespoke model had an AUC of 0.748, compared with 0.741 for OGP1. For 26–30 year olds, the AUC of 0.793 for the bespoke model is just 0.004 above that of OGP1. Among those aged 51+, many of the OGP1 dynamic items were absent from the bespoke model, but its AUC of 0.785 was again only moderately higher than the 0.780 of OGP1. It is concluded that risk factors for non-violent reoffending do not differ greatly across age groups.

Figure 3.9: Non-violent reoffending within 24 months of sentence/discharge, by grouped OGP score and age group

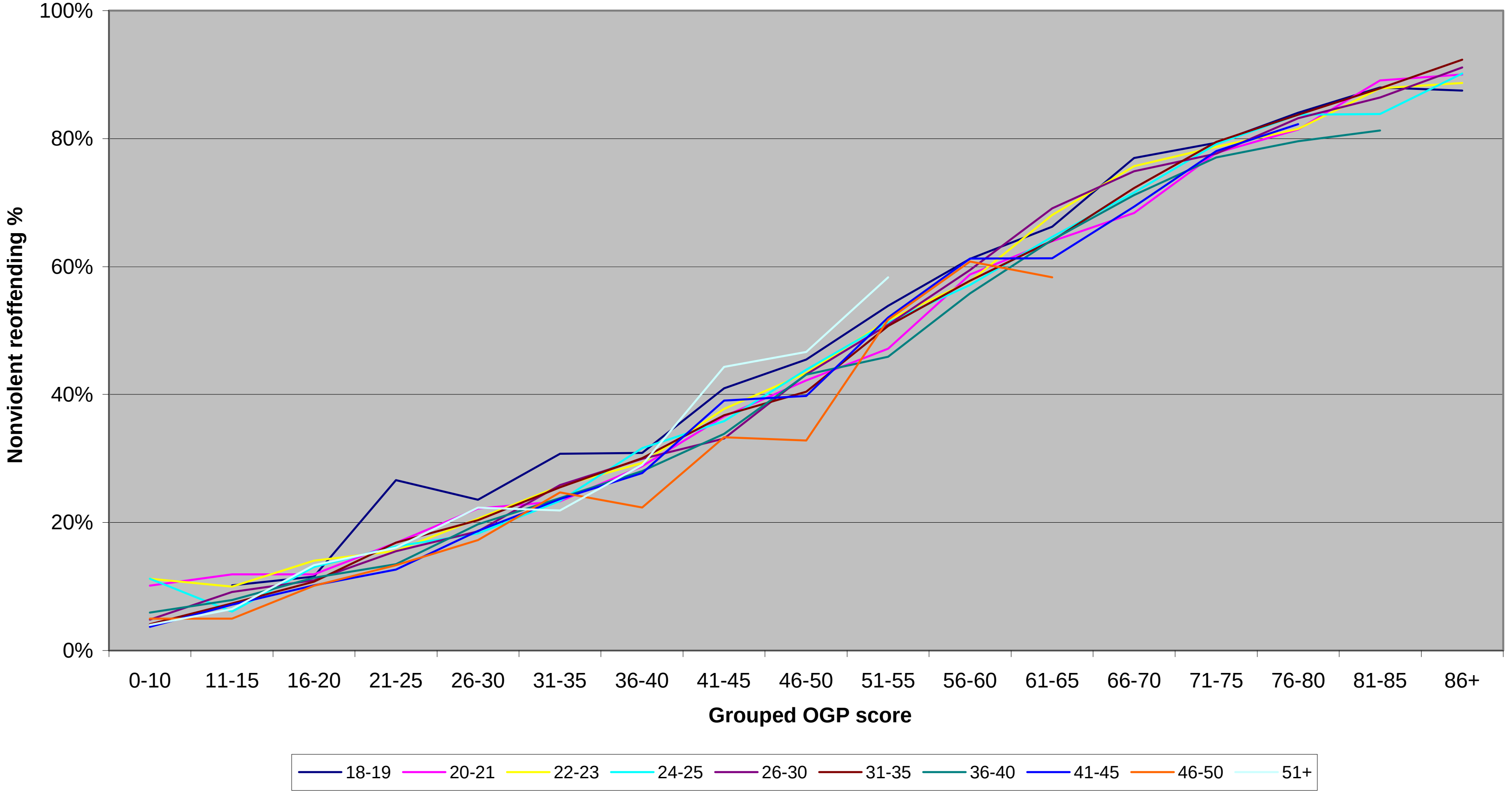
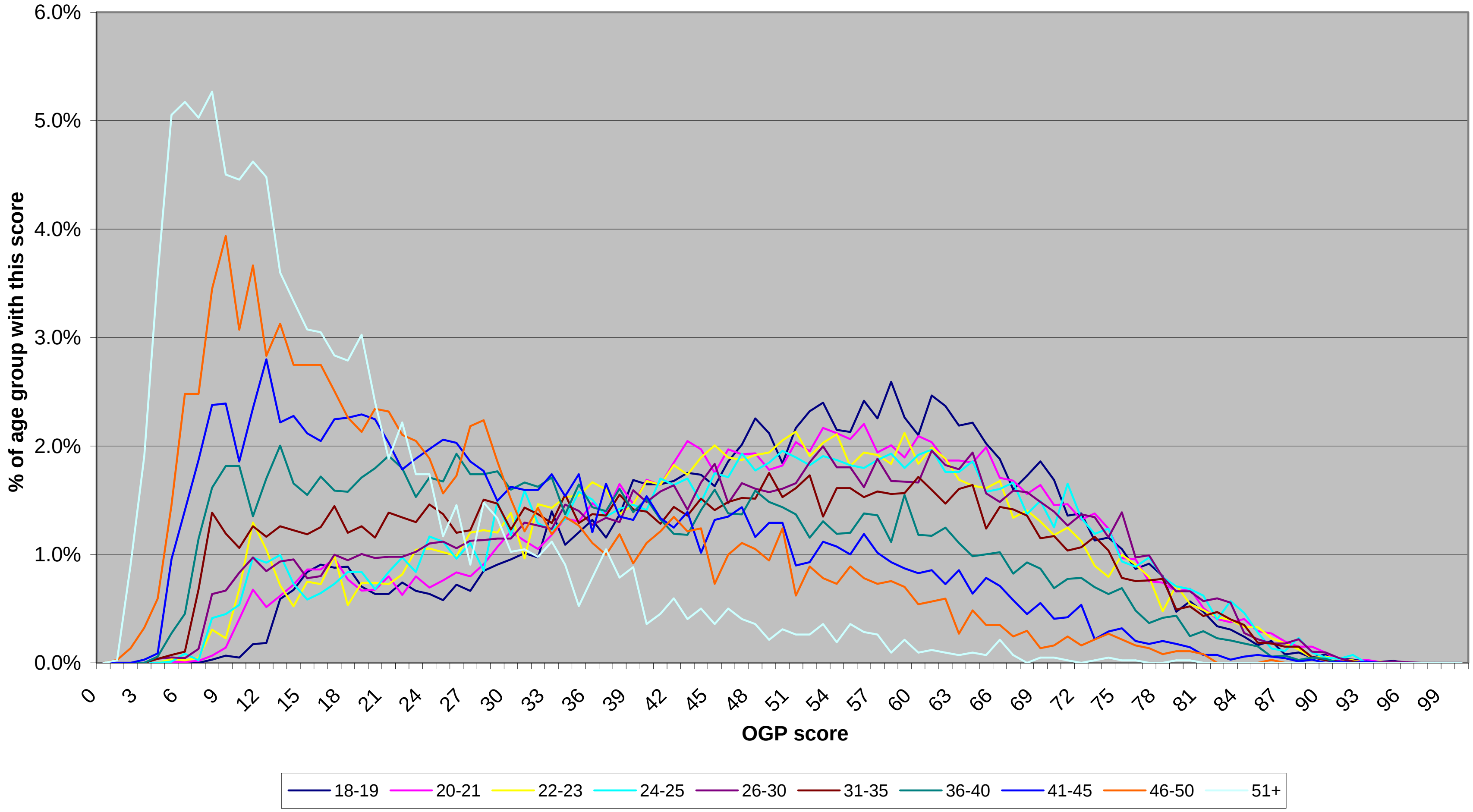


Figure 3.10: Distribution of OGP scores by age group



Absolute and relative predictive validity by age: violent reoffending

Figure 3.11 presents actual violent reoffending rates by age and OVP1 score band, while Figure 3.12 illustrates the distribution of OVP1 scores by age.

Note that OVP1 directly scores age as a risk factor, the score ranging from 0 for age 51+ to 20 for age 18–19. The very youngest offenders are considered most likely to reoffend: the score of 10 is reached by age 26–30. This direct scoring of age is the cause for the lack of results in Figure 3.11 for scores of 0–20 for age 18–21, 0–15 for age 22–23, 0–10 for age 24–25, and 0–5 for age 26–40. This also affects the top end of the score range, but is less initially obvious as the tails of these distributions are in any case very long, as very few offenders score towards the top of the 'violent sanctions' or total dynamic score ranges. The top 1% of offenders aged 18–19 scored 77+, whereas the top 1% of offenders aged 26–30 scored 67+ and of those aged 51+, the top 1% scored 58+. The top 5% of these groups scored 70+, 59+ and 48+ respectively. These are far below the highest possible scores (100, 90 and 80 respectively), but the extent to which they fall short seems consistent across age groups.

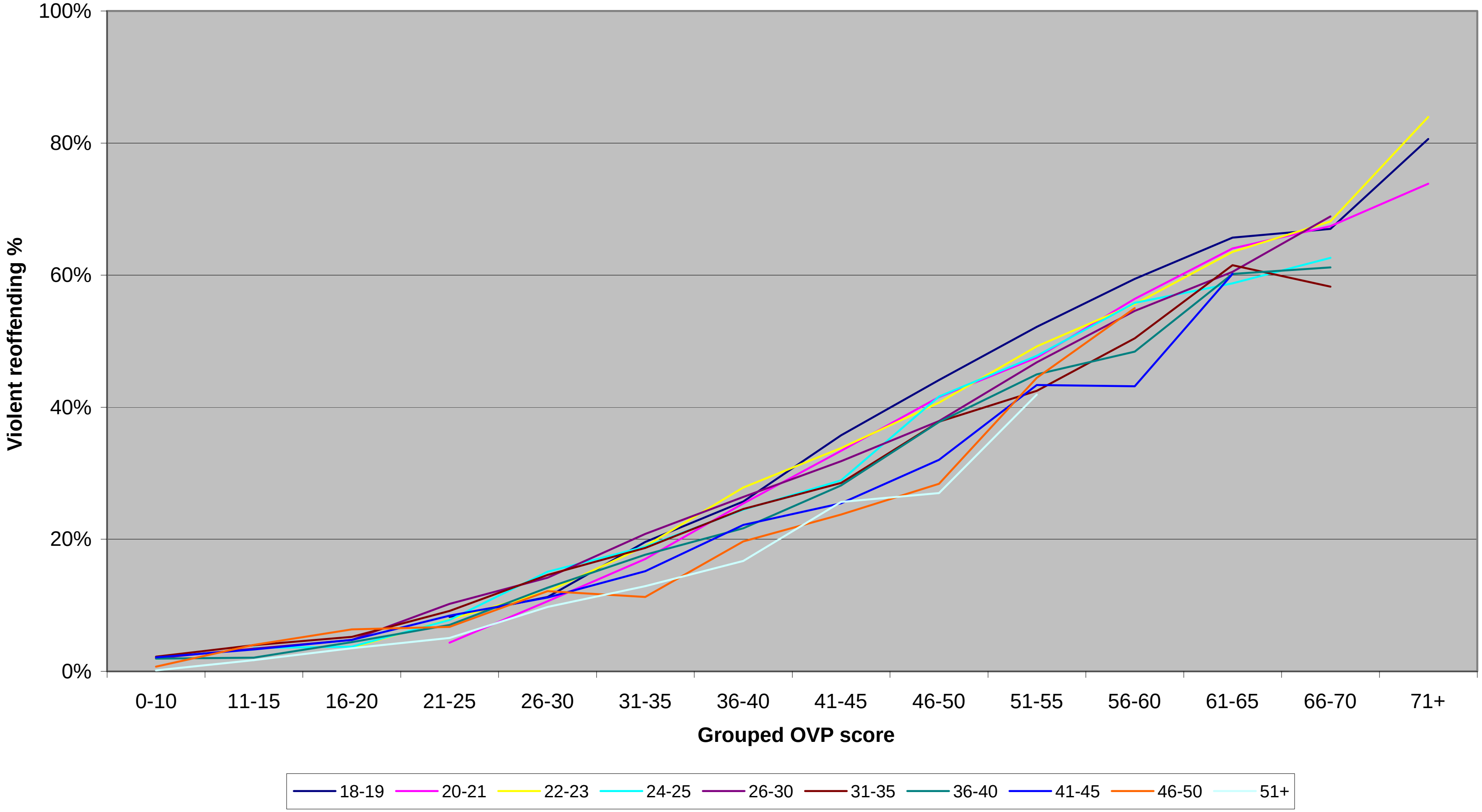
OVP1 appears to underestimate differences in reoffending rates between age groups within several score bands. At each band between 36 and 60, a difference of at least ten percentage points existed between one of the three youngest age groups (18–23) and one of the three oldest age groups (41+), with some such differences being as large as 15 percentage points and almost all being over five percentage points.

These results may imply that OVP1 gives insufficient weight to age as a predictor, and a logistic regression model confirms that prediction in the current sample would improve slightly if age was more strongly weighted. The weight of the age factor was indeed slightly reduced during OVP1's construction, as part of efforts to make the predictor more user-friendly and dynamic, but this adjustment was only from 23 to 20 points of OVP1's 100-point weighted maximum score, and the package of adjustments of which it was part made very little difference to overall predictive validity (Howard, 2009). Differences of 10% are well beyond what can be explained by a three-point shift in OVP1 scores, which would only introduce errors in reoffending predictions of a maximum of 4–5 percentage points (less at low total scores, and ignoring the redistribution of the three points to other predictive risk factors, which should reduce the net error). The new model suggests a maximum weight of 26, rather than 23, points for age.

In searching for an explanation for the remaining age-related differences, various logistic regression models were fitted. These suggested that some risk factors (being male, having many non-violent criminal sanctions, antisocial attitudes) are more strongly associated with violence risk at younger ages, but poor temper control is more strongly associated among older offenders. The interaction between age and dynamic risk factors appears to be complex, and modelling them does not result in large improvements in predictive validity. A simple model which allows the total dynamic risk factor score to interact with age (i.e. so that dynamic risk factors assume a greater or lesser role in prediction depending on the offender's age) does not improve predictive validity. An alternate explanation is that changes in violent reoffending patterns between 2002–04 and 2004–06 has increased the strength of the association between age and violence. A degree of random variation between the original (2002–04) OVP1 sample and the current sample may also have occurred.

Standardised AUCs show that the shape of each age group's risk distributions accounted for OVP1's greater predictive validity at higher ages. When the distributions of 26–30 year olds' scores was applied to 18–19 year olds, an AUC of 0.738 resulted, and when applied to those aged 51+, an AUC of 0.719 was found. These compare with 0.702 and 0.806 respectively for the unstandardised AUCs. The fall in AUC for the oldest offenders is likely to be due to the elimination from study of this group's high proportion of low scoring offenders. In other words, OVP1 identifies likely violent reoffenders well among older offenders because most of this group have a very low probability of future violence. Differences in violent reoffending rates between those with low-medium to medium-high scores are actually greater among the youngest offenders.

Figure 3.11: Violent reoffending within 24 months of sentence/discharge, by grouped OVP score and age group



3.4 Implications

This chapter presents several important results. The sample used comprised offenders starting community sentences or post-custodial supervision in 2004–06. Their overall violent and, especially, non-violent reoffending rates were less than predicted by OVP1 and OGP1 respectively, which were developed using data on similar offenders from 2002–04. These results, including a greater reduction for female than male offenders, are consistent with reductions in proven reoffending recorded in official statistics (Ministry of Justice, 2013a).

However, the non-violent reoffending rates of Asian and Black offenders were greater than predicted, as were the non-violent and violent reoffending rates of offenders aged 18–19. Within offender sub-groups, the relative predictive validity of both OGP1 and OVP1 was greater for White offenders, female offenders, and those in the older age groups. Calculations of standardised AUCs demonstrate that these gender and age differences are primarily a result of different distributions in underlying population risk – the predictors actually have fairly similar levels of effectiveness at distinguishing reoffending probabilities for individuals within those different demographic groups.

Gender differences in predictive validity have not been studied widely: Coid *et al.* (2009) found that most risk assessment instruments had greater predictive validity for male prisoners than female prisoners, but these differences were not significant. Rettinger and Andrews (2010) found that gender-neutral risk factors from the LSI-R risk assessment system predicted well for adult females, with little evidence that proposed gender-specific risk factors had incremental validity; they suggest that it is feasible, though not proven, that gender-specific issues may affect responsivity (e.g. delivery of interventions to female offenders through a strong therapeutic alliance with staff).

The most sustained differences are by ethnicity, with both predictors working less well for Black offenders and OGP1 also working less well for offenders of Mixed ethnicity. Literature comparing predictive validity between the ethnic groups as categorised here could not be identified. For example, the one relevant report published by the Correctional Service of Canada compares aboriginal and non-aboriginal populations (Sioui, Thibault & Conseil, 2002), rather than involving BME populations with ethnic origins outside Canada. This study did however corroborate the findings in this chapter through its finding that risk factor and level of need variables were less predictive among the minority (aboriginal) population.

A number of possible explanations can be advanced for the weaker relative predictive validity of OGP1 and OVP1, and the higher reoffending rates observed after controlling for predicted rate, among BME offenders. One possibility is that true differences in offending behaviour exist, with greater levels of reoffending among some BME offenders with given dynamic and static risk factors; this explanation would explain higher rates but not weaker relative predictive validity. (A subset of this explanation, for dynamic risk factors, might be overlenient assessment of BME offenders' risk factors by cautious probation staff.) Another possibility is that bias in the criminal justice system leads to greater prosecution of BME groups than those of White ethnicity. These explanations are speculative – it is not possible to test them using OASys and PNC data alone. However, evidence on how individuals of differing ethnicities are treated by the criminal justice system has been summarised elsewhere (Ministry of Justice, 2011).

In terms of the development of the next iterations of the predictors (see Chapters 8 and 9), the logistic regression analyses run here suggest that there is limited scope to improve the validity of OASys-based predictive scores for BME offenders by building separate scores for each ethnic group. These analyses achieved limited increases in predictive validity, and the relatively small absolute number of cases in each BME group presents the risk of generating new models which are either too cautious (i.e. do not include all possible risk factors, thereby reducing predictive validity) or not cautious enough (i.e. that work well in the sample they are developed on, by capitalising on chance variation, but work less well in the future). Ethnicity is not currently included in the predictors, and the residuals found here are too small to provide statistical justification for the controversial step of including ethnicity as a risk factor in a model to be used with all offenders. The potential for separate models for female offenders or particular age groups also seems limited.

For NOMS to produce separate models or include interactions in future revisions of OGP or OVP, there must be a theoretical underpinning for diversity-related differences in risk factors, and it must be possible to adjust for any such proven differences in a way which can be understood and therefore correctly interpreted by assessors. The general conclusion is that the overall predictive validity of OGP1 and OVP1 remains good. The two predictors are valid for all offenders, with most of the risk factors in OGP1 and OVP1 being valid for each sub-group, and there is currently no indication that alternate means of risk assessment would provide a meaningful improvement. However, future revisions of these predictors and the OGRS predictor should refit the age and gender terms and therefore correct for over- and under-prediction. The construction and validation of these next-generation tools should consider the modelling of age and gender carefully, and clearly present any concerns relating to validity by age, gender and ethnicity so that offender managers and other stakeholders are fully aware of how the predictors should be used responsibly.

4. Measuring changes in likelihood of reoffending

The study reported in this chapter examined whether scores on supposedly dynamic risk factors changed over the course of probation supervision, and whether changes in risk factor and predictor scores were associated with changes in reoffending risk. Key points are as follows:

- Mean OGP1 and OVP1 scores fell over the course of offenders' supervision. Scores fell more for non-reoffenders than reoffenders, even though non-reoffenders had lower initial scores. Accommodation, drug misuse and alcohol misuse scores were especially dynamic, with the greatest net reduction being in alcohol misuse, though two OVP1 risk factors did not demonstrate dynamic properties.
- Prediction of reoffending was improved by accounting for changes in dynamic risk, by using current rather than initial assessments. Changes in most OGP1/OVP1 risk factors contributed incrementally to the prediction of reoffending.
- These findings demonstrate the value of reviewing OASys assessments during probation supervision. Reviewing assessments improves prediction of reoffending by keeping dynamic risk factors up to date, and offers an evidence-based mechanism for gradual reductions in the resources allocated to a case. When designing the next iterations of the reoffending predictors, a methodology should be used which accounts for changes in dynamic risk factor scores.

4.1 Context

The value of measuring purportedly dynamic risk (or protective) factors has been questioned by some researchers. Ideally, for the purpose of prediction itself and for interventions to reduce offenders' risk levels, risk factors would be causal: to be causal, risk factors must be capable of changing (i.e. genuinely dynamic) and must be associated with changes in the likelihood of recidivism when they do change (Kraemer *et al*., 1997). On a practical level, it is easier to justify the commitment of scarce practitioner time to the assessment of dynamic risk factors if doing so improves the assessment tool's predictive validity. Some recent empirical evidence has suggested that utilising both static and dynamic risk factors promotes greater predictive validity than utilising only one or the other (Andrews and Bonta, 2007; Kroner *et al*., 2007), but other findings have suggested only very limited benefits (Campbell, French and Gendreau, 2007; Hanson and Morton-Bourgon, 2009). Douglas and Skeem (2005) observed that the lack of evidence to date supporting the use of dynamic risk factors in actuarial scales may simply stem from a failure to use genuinely dynamic measures in such research. For example, the prominent actuarial scale Violence Risk Assessment Guide (VRAG; Quinsey *et al*., 1998) measures parental factors and alcohol misuse using lifetime or wholly historic measures which are effectively static. Some research has shown that prediction of recidivism is improved by repeatedly measuring dynamic risk factors over the course of supervision, but these studies (most recently Jones, Brown and Zamble, 2010) have been small-scale and have used North American data.

Examining the nature of changes in OGP1 and OVP1 scores over time, and whether they are predictive of changes in likelihood of reoffending, will contribute substantially to the research literature on the relative merits of static only and static/dynamic actuarial risk prediction (see also Howard and Dixon, 2013), and provides several opportunities to NOMS. Evidence that changes are indeed related causally to reoffending would support the use of OASys reviews as opportunities to revise resource allocation and Risk of Serious Harm (RoSH) classifications. On the other hand, failure to find such causal relationships would suggest that OASys reviews are less valuable than had been supposed. If changes in the total OGP1 and OVP1 scores demonstrate causal association, then more detailed analysis may indicate which dynamic risk factors have the strongest causal associations with reoffending, as well as which factors change most often. Focusing interventions and supervision upon factors which are both strongly dynamic and strongly causal would help to maximise reductions in reoffending.

The aims of the study were therefore to:

1. Measure the extent of change in OGP1 and OVP1 scores.
2. Estimate the overall impact of changes in OGP1 and OVP1 scores and the passage of time on the predictive validity of these scores.
3. Improve understanding of the extent to which individual risk factors are causally dynamic, thus representing promising treatment targets.

4.2 Approach

Sample

OASys assessments completed by 31 March 2008 were filtered to select those completed at the start of a community sentence or at discharge from custody, or at Pre-Sentence Report (PSR) stage – the assessment needed to have been completed within three months of the sentence or discharge date. The assessments were further filtered to remove those with missing dynamic risk factor or RoSH data, missing offender demographic data, or missing/inconsistent sentence data. It was also ensured that there was only one assessment for each offender's period of contact with the service. These assessments were submitted to the Ministry of Justice's (MoJ's) Police National Computer (PNC) research database. Following matching, 221,157 assessments remained for inclusion in at least some survival analyses – in all these cases, OGRS3, OGP1 and OVP1 scores could be calculated and the sentence/licence length was at least four months.

The follow-up started on the day of an offender's conviction leading to a community sentence or upon discharge from custody for their index offence, and it continued until either the offender committed the offence type being studied or a ***censoring*** event occurred. Censoring occurred when the offender:

(i) reached the ***cutoff date*** for a reliable PNC follow-up without reoffending;
(ii) was imprisoned for any offence;

(iii) was recalled to custody;[42] or

(iv) received a further OASys ‘start’ or PSR assessment.

The cutoff date for this study was 2 July 2009. When data were drawn from the PNC database, it had last been synchronised with the operational database on 2 July 2010; thus allowing dates of reoffending and follow-up periods until a year previously, as an offence committed after this date would too often have not yet resulted in a PNC-recorded conviction. Follow-up periods were therefore ‘censored’ (cut off) at this point, if imprisonment, or a further OASys start or PSR assessment had not censored them at an earlier point. For imprisonment, the analysis used the date of sentence, but for reoffending it was the date of offence. The use of sentence date for imprisonments meant that offenders were only removed from the follow-up at the point at which it was clear that they were no longer at risk of committing further offences in the community. Imprisonment could either be for a new offence not under study[43] (e.g. a non-violent reoffence, when violent reoffending was the outcome of interest) or for a pseudoreconviction (i.e. an offence of any type committed before the start of the follow-up period but brought to justice afterwards).

Clause (iv) above was included as the offender’s static factors would have been rescored when a new ‘start’ or PSR assessment occurred; it also guarded against double counting of assessments.[44] In this study, follow-up periods ranged from one day to over six years. However, the longest follow-ups were rare, as very few offenders were assessed with the electronic version of OASys before 2004, and therefore calculations ceased at the five year point to provide clarity when presenting and interpreting the results.

Relatively small numbers of offenders (544 for any reoffending) were excluded from one or more analyses because they committed the reoffence of interest or were recalled to custody on ‘day zero’ of the follow-up – plainly, these offenders were not of interest when studying the impact of changes in risk assessment score as community supervision progresses. The numbers of cases included in various tables in the results section vary accordingly.

Offenders could be included as multiple cases when they were subject to separate sentences. The mean follow-up length was 3.1 years, with the cases dropping off as follows:

- 172,354 (78%) could be followed up for 12 months;

[42] The use of recall data as a source of censoring information was innovative, and checks confirmed that, by removing offenders with limited opportunities to reoffend, it improved predictive validity.

[43] It will not be for the offence under study because, as described above, the outcome of interest in a survival analysis is the date of reoffending, whereas censoring only occurs on the date on which a custodial sentence is passed. (No reliable information on remand periods is unavailable.) Therefore, a reoffence cannot be discounted through censoring by its own custodial sentence, as the offence date precedes the imprisonment date. However, if (say) this offence was non-violent then while overall and non-violent follow-ups would be uncensored, a follow-up for violent reoffending would be censored unless a separate act of violent reoffending occurred before the non-violent reoffence’s imprisonment date.

[44] Start of order/licence assessments are not counted as further assessments for clause (iv) when the index assessment was the PSR for the same order. A qualifying further assessment would, if it includes adequately complete data, lead to inclusion as a separate case in our sample; failure to censor at this point would therefore allow double counting. Moreover, at this point the static factors in an assessment are likely to be rescored.

- 112,203 (51%) for 24 months;
- 65,054 (29%) for 36 months;
- 34,220 (15%) for 48 months; and
- 10,796 (5%) for 60 months.

Among the whole sample, 87% were male, while 17% were aged 18–20, 19% aged 21–24, 47% aged 25–40 and 17% aged 41+. They included 24% on licence from a custodial sentence, and 21% domestic violence perpetrators. Principal current offences were violent for 39% of cases and sexual for 2%. While the level of attrition reported above illustrates a risk inherent in using operational data, the sample still appears to have been representative of OASys-assessed offenders.[45]

Procedure

OASys assessments were linked for all contact periods – continuous periods of contact between NOMS and an offender while under supervision in the community – using a combination of name, date of birth, gender and sentence details. In a contact period, the OGP1 and OVP1 scores at the initial assessment (i.e. at the start of community supervision) were copied across all assessments. The OGP1 and OVP1 scores at the current assessment were also calculated, together with the changes in score since the initial assessment and since the previous assessment. Similar calculations were done for all dynamic risk factor components of the OGP1 and OVP1 scores; the static factors were calculated from PNC data and fixed at their initial scores throughout. (The standard approach for static tools, e.g. OGRS3, was followed by not recalculating static risk factor scores as offenders aged during the follow-up.)

Changes in risk factor scores were accounted for in Cox regressions by ***time-dependent covariates***. These are covariates whose values change over the course of the follow-up, i.e. because the offender has had a new OASys assessment. These covariates were handled by treating the reassessment as an additional form of censoring, while allowing the follow-up to be split into the periods between OASys assessments. For example, if the offender was reassessed after 90 days and reoffended after 120 days, they were included twice in the Cox sample: once for the 0 to 90 day period,[46] with the risk scores from their initial assessment, and once for the 90 to 120 day period, with the risk scores from their reassessment.

The predictive validity of the risk prediction scores was measured using the ***Concordance Index (C)*** (Harrell, Lee and Mark, 1996). Dealing with the combination of time-dependent covariates and C measurement required an innovative measurement approach. Where risk predictor scores could vary

---

45 The data completeness filters had little effect on the sample's characteristics. Among all OASys start and PSR assessments completed by 31 March 2008, 86% were male, 17% aged 18-20, 19% aged 21-24, 47% aged 25-40, 16% aged 41+, 23% of those with recorded sentences were on licence, 21% were perpetrators, 40% violent and 3% sexual. Note that the 'violent' offences are all those classified as such by OVP1.

46 Note that all our 'day' counts are relative to the individual offender. For an offender whose follow-up started on 1 March 2005, 'day 90' is 29 May 2005, whereas for one whose follow-up started on 1 March 2006, 'day 90' is 29 May 2006.

over time, C was calculated by making comparisons between each reoffender and each offender who had survived to that day but not yet reoffended, using the scores in effect on that day. This simulated the ability of the predictor to highlight the highest-risk offenders at the points in the follow-up when reoffending actually occurred.[47] For consistency, this approach was also used to calculate C when the predictor scores were held constant at their initial follow-up values, and for static predictors (i.e. the OGRS3 score).[48]

The day-by-day nature of C calculation meant that daily C scores could be summed and weighted to produce C scores for intermediate periods, measuring the ability of the predictor to distinguish between reoffenders and non-reoffenders among those still in the follow-up at that point in time. This chapter includes some results for one- and four-month periods.[49]

A further adjustment to the calculation of C was made to account for the varying dates of the start of follow-up. As some offenders would be censored at a 16 month cutoff date while others would have potential follow-ups of five years, the calculation of C would have been slanted towards the earlier months where no/few offenders would be artificially censored in this way. Weights based on follow-up start dates were therefore applied to the C calculation to multiply up the numbers of comparisons in later months to the numbers which would have been made if all offenders had started follow-up by 2 July 2004 (i.e. had been potentially at risk for five years or more). Weights ranged from 1 (months 1 to 15) to 12.4 (month 60). This calculation adjustment did not redress any of the legitimate bias towards the earlier months caused by the other causes of censoring: reimprisonment, new supervision episodes, and recall to custody. In practice, the adjustment had little effect on final C scores.

PNC data were processed to determine dates of reoffending for:

(i) all offences;

(ii) OGP1-class offences; and

(iii) OVP1-class offences.[50]

47 So, assume a study with five offenders labelled V to Z, of whom V had the shortest follow-up through to Z having the longest, where W and Y were reoffenders whose first reoffences occured on W-day and Y-day respectively. W's score on W-day could be compared with the scores of X, Y and Z on W-day, and Y's score on Y-day with Z's score on Y-day. (Both Y and Z may have changed score between W-day and Y-day.)

48 Note that confidence intervals for C can conventionally only be calculated through resampling methods such as the bootstrap (Harrell *et al.*, 1996); for very large samples such as those used in this study, this imposes impractical computational demands, and thus no confidence intervals or standard errors can be provided. Attempts to circumvent these problems in medical statistics (Raykar *et al*., 2008) are also highly involved yet produce narrow confidence intervals on samples in the low hundreds, suggesting that intervals would in any case be extremely narrow with our very large samples, allowing us to assert with some certainty that differences in C scores are due to real differences in predictive validity.

49 In the above example, if offender W reoffended in day 17, and Y sometime after day 31, then both the day 17 and month 1 score would be based on the comparisons between W and offenders X, Y and Z.

50 Dates were also calculated for homicide and wounding reoffending so that further checks could be conducted for the most serious non-sexual violent offences.

Censoring dates were calculated for each offender, and individual valid follow-up periods were therefore determined. The actual number of pre-censoring assessments completed for each offender was constant when considering different types of reoffending, but the number of pre-censoring ***or*** reoffending assessments varied according to the reoffence type.

Table 4.1 reports the numbers of assessments followed through successive four-month periods[51] and within each of the first four months. Four-month periods were chosen as this was the required frequency of OASys reviews in the community (for the years being studied), as set out in the 2007 National Standards (Ministry of Justice, 2007). The table sets out the numbers who were censored, the numbers who had reoffended and the proportions of surviving, non-reoffending offenders with any post-initial assessment and with any change in any OASys dynamic risk factors.[52] Howard and Moore (2009) showed that a substantial proportion of offenders never change on any dynamic risk factors, concluding that a significant fraction of these must be due to the 'cloning' of previous assessments rather than a genuine lack of change.

Table 4.1 shows that more than five in every six offenders eventually had at least one reassessment, and more than three-fifths had at least one score change, allowing for censorship of follow-ups. Reassessment and change peaked in the fourth month – recall that the 2007 National Standards stated that reassessment should occur within four months – with relatively little activity in the first two months.[53] Most of those with reassessments experienced their first reassessment within a year, and almost all within two years; the same was true of score changes. This is unsurprising: as well as the influence of National Standards, many Community Orders last for two years or less, and all post-custody supervision of those with sentences of four years or less (and some of those on longer sentences, if they were remanded before sentence) will last for two years or less.[54] Therefore, beyond the two-year point, most offenders would no longer have been eligible for reassessment.

---

51 In this table only, for illustrative purposes, survival calculations were made on a four-monthly rather than daily basis. An offender censored during a four-month period was not eligible to be counted in the measurement of reoffending, reassessment or change in dynamic risk factors during that period. Imprisonment other than for pseudoreconvictions cannot and should not be counted, as such cases would involve genuine reoffending leading to imprisonment within the same four-month period. New start/PSR occasions and recalls were noted when they preceded (on a day-by-day basis) any reoffending.

52 This included any item counted towards section scores in OASys prior to its August 2009 revision, plus the OVP1-scored items 2.6 (acknowledges impact of offending) and 10.7 (current/pending psychiatric treatment). We used the fuller set of questions preceding August 2009 in order to ensure that minor changes were taken into account.

53 This supports the earlier exclusion of cases eligible for under four months supervision after release from custody.

54 The pattern of censoring in Table 4.1 is due to the sample start dates running until March 2008 and reoffending data being available until July 2009; cut-off date censoring therefore commences in month 16. The overall hazard of any reoffending drops steeply in the first year, and more slowly thereafter. Note that this pattern is not necessarily applicable to individual offenders or for sub-groups with particular OGP1 and OVP1 scores. The overall hazard is initially calculated from all offenders; many of the highest-risk offenders reoffend early, leaving the later hazards to be calculated from the remaining uncensored offenders whose average risk level will be lower than the initial average. Howard (2011) provides hazards for sub-groups of offenders, where such changes in sample composition should have much less extreme effects.

Table 4.1: Life table tracing reoffending for any offence and censoring over a 5-year follow-up

Time period	Number at start of period	Number censored	Number of non-censored reoffending (% hazard)	Number surviving to end of period	Cumulative % with no reassessment by end of period	Cumulative % with no change in dynamic risk factors by end of this period
0–4 months, of which	220,613	7,098	39,412 (18.5%)	174,103	51.2%	73.2%
0–1 month	*220,613*	*1,793*	*13,586 (6.2%)*	*205,234*	*95.9%*	*98.3%*
1–2 months	*205,234*	*1,556*	*10,741 (5.3%)*	*192,937*	*91.0%*	*95.8%*
2–3 months	*192,937*	*1,422*	*9,006 (4.7%)*	*182,509*	*76.8%*	*87.5%*
3–4 months	*182,509*	*1,256*	*7,150 (3.9%)*	*174,103*	*50.7%*	*71.7%*
4–8 months	174,103	3,259	20,981 (12.3%)	149,863	29.7%	56.2%
8–12 months	149,863	1,690	13,757 (9.3%)	134,416	23.8%	47.9%
12–16 months	134,416	3,806	10,262 (7.9%)	120,348	21.0%	44.2%
17–20 months	120,348	13,902	6,993 (6.6%)	99,453	19.5%	42.2%
20–24 months	99,453	13,542	5,048 (5.9%)	80,863	18.5%	40.3%
24–28 months	80,863	11,088	3,662 (5.3%)	66,113	17.9%	39.6%
28–32 months	66,113	9,318	2,728 (4.8%)	54,067	17.5%	39.2%
32–36 months	54,067	8,310	1,874 (4.1%)	43,883	17.2%	38.9%
36–40 months	43,883	6,845	1,397 (3.8%)	35,641	16.9%	38.7%
40–44 months	35,641	6,264	1,051 (3.6%)	28,326	16.7%	38.6%
44–48 months	28,326	5,364	734 (3.2%)	22,228	16.5%	38.5%
48–52 months	22,228	4,772	513 (2.9%)	16,943	16.1%	38.4%
52–56 months	16,943	5,101	311 (2.6%)	11,531	15.9%	38.4%
56–60 months	11,531	4,472	163 (2.4%)	6,896	15.5%	38.3%

Note. Reassessment and change in dynamic risk factors are only calculated for those surviving the period. The % hazard equals N reoffending / (N at start – N censored). The cumulative percentage equals (1 – period-1 %)*(1 – period-2 %)*...*(1 – current-period %). The final day of a four-month period is counted as part of that period and not the following period e.g. day 122 is part of ‘0–4 months’ not ‘4–8 months’. Similar rules apply in the single-month periods; event numbers for these months do not sum to those for ‘0–4 months’, nor do cumulative %s match, due to the interactions of censoring and reoffending rules when the four months are treated separately rather than together.

The analysis of score changes assumes that reassessments represent pure measures of change as and when they occur. This is untrue in one sense, in that changes are hidden from the view of researchers when there is no real reassessment (Howard and Moore, 2009). Where meaningful reassessment does occur, it usually conforms to a set schedule (due to the National Standards that applied at the time). In contrast, changes in circumstances will not occur conveniently on-schedule. If the practitioners have some awareness of a change in circumstances before this formal review, they will therefore know about the change before they record it. This mismatch between the timings of real changes and review assessments implies that offenders' behaviours in the periods before review will, on average, have been more like their post-review behaviours than their scores suggest. Furthermore, practitioners might act upon information about negative changes (increases in risk score) to suppress increased criminality, and therefore counteract the changes. All such effects will reduce both the measured frequency of changes in score, and (for those changes which are recorded) their measured effect on reoffending.

4.3 Results

Changes in score

Table 4.2 shows how the dynamic risk factors scored in OGP1 and OVP1 changed from one risk assessment to the next.[55] There were 393,893 further assessments in the OGP1 follow-up, and 413,060 in the OVP1 follow-up. The overall results for OGP1 and OVP1 are very similar:

- The dynamic elements of both scales had initial means slightly below 12 points and mean absolute score changes slightly below 1.3 points between any given pair of successive assessments.
- The total score changed in about two-fifths of pairs, so the mean score change will have been greater than three points in those assessments where it did change.[56] The mean change in score therefore was about 11% of the mean size of the initial dynamic score.
- In both cases, the mean net scores fell by over 0.4, so about two-thirds of the score change was negative. All mean net score changes, on both scales, were significantly different from zero (p<.001).

In OGP1, accommodation and drug misuse scores showed the greatest absolute change. 'Regular activities encourage offending' and thinking/behaviour showed the greatest net decrease in scores. Thinking and behaviour changed most often, but 'regular activities' and drug misuse did so least often: when changes in these risk factors did occur, the scoring rules (see Appendix B) made it likely (drugs) or certain ('regular activities') that scores would change by multiple points. The drug misuse scale is also notable in that most offenders always score zero, with the scores of the minority being high and prone to change. Attitudes scores showed the smallest net reduction over time.

55 These are weighted scores, and it should be noted that the underlying raw score ranges differ, as shown in Appendix A.
56 For offenders around the low/medium or medium/high boundaries, three-point changes in the 100-point score trigger changes of 3% to 5% in the likelihood of non-violent recidivism or 4% to 6% in the likelihood of violent recidivism. As Appendix B shows, a six-point change in OVP1 score results in a 10% change in likelihood of violent recidivism in over three-fifths of cases, when scores are between roughly 35 and 70.

In OVP1, offenders' statuses on the two yes/no questions – recognising the impact of offending and psychiatric treatment – seldom changed. Alcohol misuse showed the greatest mean absolute change and the greatest fall in net score of any scale from either predictor; reductions in alcohol misuse score accounted for about half of the total net fall in OVP1 score. Temper control changed less often, but changes in this score also usually indicated reductions in risk. Changes in accommodation, employability and attitudes scores were similar to those in OGP1, though not identical due to the different item weights.[57]

Table 4.2: Changes in OGP1 and OVP1 risk factors between successive assessments

Risk factor (maximum points)	Mean (SD) of weighted scores at initial assessment	Mean absolute change (% of initial mean)	Mean net change (% of initial mean)	% with any change
OGP1 risk factors				
Total score (100)	*42.6 (20.4)*	*1.29 (3%)*	*-0.44 (-1%)*	*40.7%*
OGRS3 [static] score (60)	30.9 (14.9)	n/a	n/a	n/a
Total dynamic score (40)	*11.7 (8.1)*	*1.29 (11%)*	*-0.44 (-4%)*	*40.7%*
Accommodation (5)	1.28 (1.81)	0.34 (22%)	-0.06 (-4%)	13.7%
Employability (5)	2.07 (1.57)	0.18 (9%)	-0.05 (-3%)	12.9%
Regular activities encourage offending (5)	1.91 (1.97)	0.24 (13%)	-0.11 (-6%)	8.2%
Drug misuse (15)	2.74 (3.91)	0.31 (11%)	-0.09 (-3%)	8.5%
Thinking and behaviour (5)	2.39 (1.37)	0.24 (10%)	-0.12 (-5%)	16.6%
Attitudes (5)	1.28 (1.12)	0.16 (12%)	-0.02 (-2%)	12.9%
OVP1 risk factors				
Total score (100)	*39.4 (13.8)*	*1.26 (3%)*	*-0.41 (-1%)*	*38.6%*
Static score (60)	27.6 (9.3)	n/a	n/a	n/a
Total dynamic score (40)	*11.7 (7.3)*	*1.26 (11%)*	*-0.41 (-4%)*	*38.6%*
Recognises impact of offending on victim/community/society (4)	0.84 (1.63)	0.07 (8%)	-0.01 (-1%)	1.7%
Accommodation (4)	1.09 (1.45)	0.28 (25%)	-0.05 (-4%)	13.6%
Employability (6)	2.73 (2.00)	0.23 (8%)	-0.06 (-2%)	13.7%
Alcohol misuse (10)	3.27 (3.83)	0.45 (14%)	-0.20 (-6%)	11.5%
Psychiatric treatment current/pending (4)	0.21 (0.89)	0.02 (10%)	0.00 (2%)	0.5%
Temper control (6)	2.04 (2.28)	0.20 (10%)	-0.08 (-4%)	6.2%
Attitudes (6)	1.55 (1.33)	0.20 (13%)	-0.02 (-1%)	15.2%

Note: Initial assessment N=220,793 for OGP1, 220,997 for OVP1. Change assessment N=393,893 for OGP1, 413,060 for OVP1. All mean net changes were significantly different from zero at $p<.0001$.

[57] The accommodation section of OASys contains a special scoring rule: those scoring 2 on item 3.3 (indicating no fixed abode or transient accommodation) automatically score 2 on the other three items. It is therefore possible that offenders could suddenly move from being in an unproblematic housing situation to unexpectedly losing their accommodation, causing their weighted scores to change from 0 to 4 (OGP1) and 5 (OVP1), or moving from homelessness into a good housing situation for the opposite scoring effect. In fact, this seldom happened. Most offenders scoring highly on accommodation questions switched between no fixed abode status and other severely problematic housing situations, and their OGP1/OVP1 scores therefore only changed by one or two points.

Changes in score by final assessment of follow-up, for reoffenders and non-reoffenders

Table 4.3 refers to the most recent assessment for each case, separating cases where only one assessment occurred from those where there were multiple assessments. It therefore allows examination of how offenders' scores changed over the course of the supervision and follow-up period, among cases where there may be a score change. It summarises initial scores and score changes since the initial assessment for (i) follow-ups resulting in reoffending and (ii) follow-ups resulting in censoring (which may be for any of the four reasons set out in Section 4.2 above).

For all three types of reoffence, offenders' scores tended to reduce over time, but mean reductions in score were greater among non-reoffenders, even though the non-reoffenders had lower initial scores and therefore less capacity to reduce their scores. While these differences in initial score were considerable for those with multiple assessments, they were greater still for those with a single assessment. The high scores of reoffenders with a single assessment naturally reflect the tendency of high-scoring offenders to reoffend quickly (Howard, 2011), while the low scores of non-reoffenders with a single assessment suggests some tendency among assessors to deprioritise those with low initial scores.[58] All score reductions were statistically significant at p<.001.

Table 4.3: Initial scores and changes in score by final assessment for reoffenders and non-reoffenders

Reoffence type and predictor score used, and follow-up type and outcome	N	Initial score			Change from initial to final score		
		Mean	SD	SE	Mean *(% of mean initial)*	SD	SE
All reoffending (OGP1 score)							
Follow-ups involving one assessment only							
No reoffending	20,638	36.2	21.0	0.15	n/a	n/a	n/a
Reoffending	53,557	54.7	17.9	0.07	n/a	n/a	n/a
Follow-ups involving multiple assessments							
No reoffending	87,551	32.8	18.1	0.06	-1.49 *(-4.5%)*	3.93	0.01
Reoffending	58,867	48.3	17.6	0.07	-0.85 *(-1.8%)*	4.04	0.02
Non-violent reoffending (OGP1 score)							
Follow-ups involving one assessment only							
No reoffending	25,936	38.4	20.9	0.13	n/a	n/a	n/a
Reoffending	39,074	57.4	17.3	0.09	n/a	n/a	n/a
Follow-ups involving multiple assessments							
No reoffending	113,377	35.2	18.3	0.05	-1.23 *(-3.5%)*	4.15	0.01
Reoffending	42,406	51.2	17.3	0.08	-0.80 *(-1.6%)*	4.23	0.02

58 A minority of offenders were not eligible for OASys according to current user guidance.

Reoffence type and predictor score used, and follow-up type and outcome	N	Initial score			Change from initial to final score		
		Mean	**SD**	**SE**	**Mean *(% of mean initial)***	**SD**	**SE**
Violent reoffending (OVP1 score)							
Follow-ups involving one assessment only							
No reoffending	33,748	38.1	14.2	0.08	n/a	n/a	n/a
Reoffending	24,992	48.1	12.4	0.08	n/a	n/a	n/a
Follow-ups involving multiple assessments							
No reoffending	127,410	36.3	13.2	0.04	-1.15 *(-3.2%)*	4.10	0.01
Reoffending	34,847	45.6	11.8	0.06	-0.67 *(-1.5%)*	4.04	0.02

Validity of initial and revised risk predictor scores

Table 4.4 reports the predictive validity of OGP1 as a predictor of all and non-violent reoffending, and of OVP1 as a predictor of violent reoffending. OGRS3 is included as a comparator; as OGRS3 scores are based on static risk factors, they are not recalculated during supervision. Concordance Indices (C) are reported for the original scores – see Appendix C for the distribution of initial OGP1 and OVP1 scores – and for the scores at time of reoffending. The Indices were then calculated twice. The first included predictor score comparisons for all reoffenders; the second used only comparisons for those reoffending after at least four months of follow-up. Given the primary interest in the impact of score changes, this latter calculation removed offenders who reoffended too quickly to be expected to have a reassessment. That is, it focused on offenders where OASys review was a strong possibility. However, the former calculation has value in that it provides a global estimate of the predictive impact of reassessment without any pro-reassessment sample selection process.

Table 4.4: Predictive validity of OGRS3 and initial and current OGP1 and OVP1 scores

Follow-up range and reoffending outcome	Concordance Index (C) by predictor and scoring method					
	OGRS3	OGP1		OVP1		
		Initial	Current	Static-only	Initial	Current
All offenders						
Any	0.7147	0.7211	0.7232	n/a	n/a	n/a
Non-violent	0.7399	0.7498	0.7519	n/a	n/a	n/a
Violent	0.6824	n/a	n/a	0.7048	0.7155	0.7197
Excluding those reoffending/censored in first four months						
Any	0.7013	0.7056	0.7083	n/a	n/a	n/a
Non-violent	0.7240	0.7314	0.7342	n/a	n/a	n/a
Violent	0.6796	n/a	n/a	0.7027	0.7086	0.7136

Note. OGP1 is not designed to predict violent outcomes, nor OVP1 to predict overall ('any') or non-violent outcomes. OGP1 uses OGRS3 to score static risk factors, whereas OVP1 has a separate static score component.

All reoffending and non-violent reoffending

Table 4.4 confirms the result of Howard (2009), that OGP1 produced a modest improvement in predictive validity compared with OGRS3, for both all and non-violent reoffending. Among all offenders, the initial score raised C for all reoffending by 64 base points and the current score raised C by 85 base points. When only those with four months' follow-up were considered, the respective improvements were 43 and 70 base points. For non-violent reoffending, the improvements were 99 and 120 base points among all offenders and 74 and 102 base points among those surviving more than four months. These improvements in C were greater than for all reoffending, reflecting the intended design of OGP1. For both outcomes, among both offender groups, using the current score improved predictive validity. For both outcomes, the improvement associated with using the initial score rather than OGRS3 was greatest among those reoffending or censored in the first four months, but the improvement associated with using the current rather than initial score was greater among those surviving four months.

The difference in predictive validity between the initial and current score emerged over the first few months, as reassessments commenced (as was shown in Table 4.2). The advantage of either OGP1 score over OGRS3 was greatest in the earliest months: in month 1, it was 165 base points. All predictors had higher C in the earlier months, when the highest-risk offenders were still part of the cohort and could therefore be compared with the lowest-risk offenders. The advantage of the current over initial score peaked in months 4 to 12, between 29 and 39 base points. It persisted through the second year but vanished during the third year, while the advantage over OGRS3 gradually diminished (it is around 100 base points or slightly higher between 5 and 24 months) and became very inconsistent after the third year.

Violent reoffending

Table 4.4 confirms that using OVP1 in full produced a reasonable improvement over only using its static risk subscale, and a very large improvement (between 300 and 600 base points) over OGRS3. Among all offenders, the initial score raised C for violent reoffending by 93 base points, and the current score by 149 base points, compared with the static score. Excluding those removed from the sample in the first four months, the equivalent improvements were 59 and 109 base points.[59] These results generally have an equivalent interpretation to the OGP1 findings.

For all violent reoffending, the advantage associated with dynamic risk prediction was far greater for OVP1 than for OGP1 in the earliest months, but declined quickly. The current score had C 353 points higher than the static score in month 1, but this fell to 243 points in month 2, was never higher than 200 points after month 5, and fell permanently below 100 points after month 12. The advantage of the current over initial score peaked between 44 and 81 points between months 4 and 24. The advantage of the initial over static score fell permanently below 100 points after month 11.

Taken together, these results emphasise the real though moderate incremental predictive value of reviewing dynamic risk assessments. The value of using dynamic risk factors at all, and of reviewing them, appears to be greatest for the most serious violent offending. Moreover, in general the value of reviewing assessments is greater for violent than non-violent reoffending. This suggests that the role of rapidly changing, or ‘acute’, dynamic risk factors may be greatest in the most serious forms of reoffending.

[59] Further checks found that the use of the current rather than initial score was especially valuable for homicide and wounding reoffending.

The decline over the course of follow-up in the additional predictive validity accounted for by dynamic factors may be due to the declining timeliness of measurement of such risk factors. As Table 4.2 showed, not all offenders were reassessed, and not all reassessments resulted in score change. Sometimes a lack of score change will have been legitimate but, as Howard and Moore (2009) showed through statistical analysis, there are too many ‘zero-change’ reviews to be accounted for reasonably. Therefore, if these risk factors are genuinely dynamic, and changes in their scores are prospectively related to reoffending, and yet the frequency with which they were reviewed diminished over the course of the follow-up, then it should be expected that they will have made a greater contribution to prediction early in the follow-up when the amount of time since the most recent (re)assessment tended to be shorter.

Cox regression models of initial scores and changes in score

Table 4.5 sets out the results of basic Cox regression models considering the initial score and change in score for each of the three outcomes, with OGP1 predicting non-violent and any reoffending, and OVP1 predicting violent reoffending. In predicting any reoffending, each point of the initial score and change in score were of equal predictive power. For non-violent and violent reoffending, a point of the initial score was slightly but significantly more predictive than a point’s change in score.[60]

Table 4.5: Cox regression models of all three reoffending outcomes

Predictor and reoffending outcome	Parameter estimates					
	Initial score			Change in score		
	Beta	SE	Hazard ratio	Beta	SE	Hazard ratio
OGP1 score						
Any	0.04068	.00016	1.042	0.04099	.00112	1.042
Non-violent	0.04773	.00019	1.049	0.04409	.0125	1.045
OVP1 score						
Violent	0.05976	.00032	1.062	0.05515	.00152	1.057

Note. Beta = effect size per point of predictor. SE = standard error of Beta. Hazard ratio = ratio of hazards for scores *x+1* and *x*.

The importance of individual risk factors in repeated dynamic risk assessment

Further Cox regression models were created to establish the effect of single-point changes in each of the OGP1 and OVP1 dynamic risk factors. These findings – see Tables D2 and D3 in Appendix D – can be combined with the results in Table 4.2 to generate a summary metric which is the product of the extent to which change occurs and the effect of each point of change. Table 4.6 sets out the predictors’ weights and the length of the risk factor scales (columns (1) and (2)), the results from Tables D2 and D3 (column (3)) and Table 4.2 (column (5)) and the necessary calculations to standardise correctly for the length of each risk factor scale (columns (4), (6) and (7)).

[60] For homicide and wounding reoffending, the change in score was more predictive than the initial score.

In OGP1, accommodation and attitudes appear to play a greater role in the dynamic prediction of non-violent reoffending than is allowed for by the risk factor weightings. Both account for 5 (12.5%) of the 40 point dynamic score, yet they account for 25% and 20% respectively of the changes in likelihood of reoffending during the follow-up. Employment is the least important factor, accounting for 6% of the changes against a 12.5% weighting, while drug misuse changes are important (25%) but do not fully justify this risk factor's high overall weighting (15 of 40; 37.5%).

In OVP1, similar results are found for accommodation and employability. Alcohol misuse, temper control and attitudes all have dynamic roles in proportion to their shares of the 40-point total dynamic score. Scores on the remaining two items, recognising the impact of offending and being in psychiatric treatment, have very little value as true dynamic risk factors. Impact scores change quite infrequently, and their changes are entirely non-predictive. While changes in psychiatric treatment status are reasonably predictive, these occur very infrequently.

Table 4.6: Acuteness of dynamic risk factors in OGP1 and OVP1

Risk factor	Weight in risk predictor (1)	Unweighted range of risk factor scale (2)	Beta (per weighted point) for changes in score (3)	Beta per unweighted point of risk factor scale (4)=(1)*(3)/(2)	Weighted mean absolute change (5)	Mean absolute change per unweighted point (6) = (5)/(1)	Product of unweighted Beta and unweighted mean absolute change		
							Per point of unweighted scale (7) = (4)*(6)	Across range of unweighted scale (8) = (7)*(2)	% of total product over unweighted ranges (9)
OGP1 risk factors for non-violent reoffending									
Accommodation	5	8	.037	.023	.34	.068	.00157	.0126	25
Employability	5	8	.016	.010	.18	.036	.00036	.0029	6
Regular activities	5	2	.030	.075	.24	.048	.00360	.0072	15
Drug misuse	15	10	.039	.059	.31	.021	.00122	.0122	25
Thinking & behaviour	5	8	.032	.020	.24	.048	.00096	.0048	10
Attitudes	5	8	.097	.061	.16	.032	.00195	.0098	20
Total	*40*							*.0495*	*100*
OVP1 risk factors for violent reoffending									
Impact	4	2	.001	.002	.07	.018	.00004	.0001	<1
Accommodation	4	8	.065	.033	.28	.070	.00231	.0185	26
Employability	6	8	.029	.022	.23	.038	.00084	.0067	9
Alcohol misuse	10	4	.046	.115	.45	.045	.00518	.0207	29
Psychiatric treatment	4	2	.048	.096	.02	.005	.00048	.0010	1
Temper control	6	2	.060	.180	.20	.033	.00594	.0119	17
Attitudes	6	8	.062	.046	.20	.038	.00152	.0122	17
Total	*40*							*.0711*	*100*

Note: impact and psychiatric treatment are recorded as binary variables in OASys, but are treated as 0/2 items here to allow parity with all other risk factors. Betas (3) are from Tables C2 and C3. Weighted mean absolute changes (5) are from Table 4.2. (9) = (8) / sum of all (8) values.

4.4 Implications

These results show that most of the 'dynamic' risk factors included in OGP1 and OVP1 are indeed dynamic. Scores on accommodation, the two substance misuse questions, temper control and antisocial attitudes varied often, and changes in these scores were related to changes in the likelihood of reoffending. Using scores produced at the most recent assessment was shown to improve predictive validity for all outcomes The modest nature of these improvements in terms of overall predictive validity must be set against: (i) the fact that much reoffending occurs early in the sentence before any review is necessary (i.e. both the costs and benefits of reviewing scores are accrued by the same sub-group of offenders); and (ii) the lack of any meaningful review or any review at all for a substantial minority of offenders (i.e. the average benefit across assessments which did have meaningful review was greater than shown here). The predictive power of each score was shown to be greatest soon after the assessment at which the score was determined.

The results in Table 4.6, which summarised the relative extent and importance of changes in dynamic risk factors, do not necessarily override the risk factor weightings used in OGP1 and OVP1. Initial scores on the dynamic items are also very important in prediction, but the 'acuteness' measures provide useful additional insights in focusing offender manager attention on the areas where change is most likely and may have the greatest effect. For example, Appendix D shows that accommodation became more predictive of all three types of reoffending when changes in score were taken into account; this is because accommodation status changed often and therefore initial accommodation scores were comparatively poor indicators of accommodation status later in the follow-up. The statistical method used to produce the forthcoming version 2 of OGP and OVP (see Chapter 9) should be selected in order to ensure that the risk factor weightings of these predictors incorporate data on both initial scores and score changes.

The results demonstrate that formal review processes can provide staff with important information which they can use to prioritise offenders with the greatest likelihood of reoffending, potentially reallocating resource from those offenders who have become less risky while under community supervision. If assessment resources need to be prioritised, the findings suggest that the benefits of reassessment are greater for violent than non-violent reoffending.

5. Predicting reoffending for rare, harmful offences

In the study reported in this chapter, patterns of reoffending for six types of rare, harmful offence were analysed in order to determine whether offenders specialise in these offences. Key points are as follows:

- Some degree of specialisation was found for all six offence types. For arson, child neglect, dangerous driving, kidnapping and racially aggravated offending, those with a history of such offending were three to four times more likely to commit further offences than those without, rising to eight times for blackmail. Offence-specific history was therefore a risk factor for all six offence types.
- Arson, kidnapping and racially aggravated offences were well predicted by OVP1. They should be included in the set of offences which OVP classes as violent.
- Child neglect offences were most likely to be committed by young women living with children, especially those with high scores on dynamic risk factors included in OGP1. The principal dynamic risk factor in OGP1 is drug misuse.
- Dangerous driving offences were most likely to be committed by young men with employability, lifestyle and impulsivity problems and histories of driving whilst disqualified or uninsured and/or dangerous driving. OGRS3 and OGP1 were reasonable predictors.
- Blackmail was an extremely rare offence, and the likelihood of such reoffending may be assessed using OGRS3 or OGP1.
- While most of these offences were relatively rare, rates of reoffending among those most at risk were sufficiently high that the possibility of such offending should be explicitly considered when conducting risk assessments in these cases.

5.1 Context

Public protection is a key aim of the National Offender Management Service (NOMS). As part of the offender assessment and management process, it is therefore important that offender managers should be able to identify those offenders most at risk of committing offences which cause serious harm. Scores on the OASys Violence Predictor v.1 (OVP1; Howard, 2009) estimate with reasonable predictive validity the offender's likelihood of proven recidivism involving a broad group of violence-related offences, and also identify those most likely to commit homicide and wounding with intent, the most serious non-sexual violent offences. Scores on the 'S' scale of Risk Matrix 2000 (RM2000; Thornton, 2007) are calculated for men with a known history of sexual offending when aged 16 or over, and have moderate predictive validity for future sexual offending (Barnett, Wakeling and Howard, 2010). Between them, OVP1 and RM2000 cover the stereotypically most high-harm offence groups: sexual and non-sexual violent offending, with OVP1 encompassing potentially serious offences including weapon possession, robbery and aggravated burglary, non-arson criminal damage endangering life and threats to kill, as well as interpersonal violence. However, no studies have been conducted to determine whether any actuarial prediction score can predict adequately a range of other offences which cause serious harm. These offences are not sexual, so no RM2000 scores are available, nor do some of their circumstances and motivating factors seem to fall easily within the 'violent-type' offences predicted by OVP1.

The offences selected for study, termed 'rare, harmful offences',[61] were as follows:[62]

- arson;
- blackmail;
- child neglect;
- dangerous driving;
- kidnapping; and
- racially aggravated offences.

Existing literature provides varying detail on the perpetrators of these six types of rare, harmful offence, and on the risk of recidivism for such offences. The aims of the analysis in this chapter were therefore to:

1. Summarise the risk/needs profiles and criminal histories of OASys-assessed offenders convicted of rare, harmful offences. This information may be useful for those designing new interventions, and allows comparisons with mainstream offender groups for whom existing interventions have been designed.
2. Establish the prevalence of reoffending, and whether the predictive scores within OASys differentiated between offenders more and less likely to reoffend, for each of the six types of rare, harmful reoffence, for
   a) offenders with a known history of such offending, and
   b) offenders with no known history of such offending.
3. Develop statistical models which might identify useful predictors of such reoffending. The content and validity of these models, combined with the information on prevalence and the validity of existing predictors, could help to inform decisions on whether there is practical value in managing those convicted for each rare, harmful offence differently from other offender groups.

[61] Although the results show that dangerous driving and racially aggravated offences are not especially rare. It is acknowledged that many offences can cause serious harm, but these offences are especially likely to do so.

[62] Offences related to terrorism could not be adequately studied using the available data. Consultation with NOMS staff responsible for the treatment and management of extremist offenders revealed that the offender group who had been convicted for offences under statutes such as the Terrorism Act 2006 overlapped little with the extremist offenders under their supervision, with many of the former group likely to have been convicted for activities which were actually relatively minor breaches of public order. As such, any results on 'terrorist' reoffending, which would have to be based upon the statutory offence codes, would not provide reliable information on future harmful extremist reoffending. A further possibility was to include abduction offences, either alone or together with kidnapping. However, OASys details of index offences found that around half of abduction offences had a sexual element or motivation (OASys questions 2.2F and 2.9), which is consistent with the current classification of abduction as a sexual offence. Kidnapping offences, on the other hand, had a sexual element or motivation in only 11% of cases.

Those with a history of each offence were studied separately from those without, as assessors would have been especially alert to the possibility of a new offence where the offender was already known to have committed such offences.

5.2 Approach

Sample

A list of offenders assessed using OASys by 31 March 2007 was created, filtering out assessments with missing dynamic risk factor data or Risk of Serious Harm (RoSH) ratings, and ensuring that each offender was only represented once during each period of contact with the criminal justice system. This list was submitted to the Ministry of Justice's (MoJ's) Police National Computer (PNC) research database in January 2010. After filtering out those whose index conviction (i.e. the conviction for which OASys was being completed) could not be identified on the PNC, those whose assessment was not within three months of their community sentence or discharge from custody, those for whom OGRS3, OGP1 or OVP1 scores could not be calculated,[63] and those whose follow-up commenced less than 36 months prior to the PNC extract date, 144,229 cases could be included in the analysis of 24-month proven reoffending outcomes. These offenders commenced community sentences or were discharged from custody between January 2002 and January 2007, with the bulk towards the end of this period due to improvements in data completeness and the advent of routine use of the electronic form of OASys. Offenders could be included more than once when these assessments were related to separate non-concurrent sentences.

The eligible sample included:[64]

- 22% on licence from a custodial sentence;
- 33% on Criminal Justice Act 2003 (CJA 2003) Community Orders;
- 9% on Suspended Sentence Orders; and
- 36% on pre-CJA 2003 community sentences.

The sample includes few offenders with custodial sentences of under 12 months or non-rehabilitative community sentences (e.g. Community Orders with only unpaid work and/or curfew requirements), as OASys is not routinely used with these offenders. Demographic details of the eligible sample are included in the results section.

[63] Due to missing date of birth or apparent convictions aged under 10, or missing data on OGP1 or OVP1 items.
[64] Checks confirmed that the data filtering process had little impact on the characteristics of the sample.

Procedure

PNC data processing and identification of offender groups

PNC output was analysed to count the number of separate occasions, prior to and including the index offence, in which the offender had been sanctioned (convicted or formally cautioned) for each of the six rare, harmful offences, and 20 other offence types. The 20 offence types were selected by breaking down the standard offence categorisation used by the Home Office and MoJ to give more opportunity to detect potential differences in associations between previous history and reoffending (e.g. the possible link between kidnapping and abduction meant that abduction was separated from other sexual offences; on similar logic, the remaining contact sex offences were separated from non-contact sex offences). All 26 offence types were mutually exclusive.

While each sanctioning occasion could only be scored once per offence type, no matter how many offences of that type were involved, multiple offence types could be scored from a single sanctioning occasion.[65] An adjustment was made for arson and racially aggravated offences, as it is possible to use OASys data to determine whether the current offence involved these elements.[66] This had relatively little effect on arson, with the mean number of sanctions rising from 0.039 to 0.040 and the proportion with any sanction rising from 3.5% to 3.6%, but far more effect on racially aggravated offences, with the mean rising from 0.025 to 0.040 and the proportion rising from 2.4% to 3.7%.[67]

Seven offender sub-groups were then created. The first six comprised offenders with at least one appearance for each of the offence types; the composition of these groups could and did overlap. The seventh group comprised those in none of the other six groups. Log-linear analysis was conducted to examine associations between membership of each group – that is, to determine whether the groups overlapped more or less than might be expected on the basis of statistical independence (i.e. where a certain amount of overlap would exist due to membership of each group being determined by separate, random processes).

---

65 The 20 other offence groups exclude the rare, dangerous offences which might usually belong within them. For example, if an offender is convicted of five assault offences and one kidnapping offence on a given day, this is counted as one violence against the person sanction and one kidnapping sanction; if they were only convicted of the kidnapping offence (which would be classified as violence against the person in most studies), this would be counted as no violence against the person sanctions and one kidnapping sanction.

66 OASys question 2.2D checks for arson as an element of the current offence. Question 2.3B checks whether "victim(s) targeted because of racial motivation or hatred of other identifiable group", and 2.9 for "racial motivation or hatred of other identifiable group [as a motivating factor]". The "other identifiable group" element does suggest that using question 2.9 in this way will incorrectly classify some offences.

67 The Crime and Disorder Act 1998 introduced racially aggravated variants of existing offences of violence, harassment, public order and criminal damage. Fieldwork conducted in 2000 found that most offences with racist elements could be charged as one of these new statutory offences. The new offences were being used widely by police, though with some geographic inconsistency, and were filtered on evidential grounds (sometimes down to the basic, not racially aggravated version of the offence) by the Crown Prosecution Service (Burney and Rose, 2002). Foster, Newburn and Souhami's (2005) review of policing since the Stephen Lawrence Inquiry found that police forces had improved opportunities for the public to report racist incidents, and their recording and investigation of such incidents. Thus, in the absence of exact data, we assume that issues such as plea bargaining and charge substitution, as identified by Burney and Rose (2002), balanced by the improved reporting and follow-up identified by Foster *et al*. (2005), affected past offences and potential reoffences to a similar degree to the current offence. It is therefore expected that our (enforced) use of previous sanction counts and reoffending measures derived from PNC offence codes will result in suboptimal prediction, as well as underestimation of the proportions of the caseload with racially motivated offending histories and reoffending.

PNC data were also processed to determine 24-month proven reoffending outcomes for the six offence types. An additional 12-month 'buffer' period was allowed for the offence to be brought to justice and PNC data entry to occur, summing to the 36-month period specified in the Sample above.

Risk/needs profiles and criminal histories

The demographic characteristics of offenders in each group, and of all offenders, were tabulated. The proportions with criminogenic needs for each offence, as measured using the revised scores developed by Moore (2009b) and implemented in August 2009, were also tabulated, with chi-square tests used to compare the first six groups with the seventh (no history of rare, harmful offences) group. Within the seventh group, those with different types of current offence were distinguished, to allow comparisons of the six groups with mainstream offender type. The mean number of previous sanctions for each of the 26 (six rare, harmful, plus 20 other) offence types was calculated for each of the seven groups, and for all offenders, to identify differences in previous criminal history.

Prediction of rare, harmful reoffending using existing scales

Proven reoffending outcomes for each of the six rare, harmful offences were compared for those with and without any previous sanctions for the relevant offence. As a summary measure of the apparent degree of specialisation in each offence, odds ratios for reoffending by offenders with previous sanction(s) for that offence versus other offenders were calculated. As well as simple odds ratios, logistic regression models were used to offset the OGRS3 score, and thus estimate the odds ratio for previous sanctions after controlling for this standard measure of static general reoffending risk.

The predictive validity of existing risk measures – OGRS3, OGP1, OVP1 and their static and dynamic subscales – was compared for each of the six rare, harmful offences, separately for those with and without previous sanctions for the offence in question (where sample sizes allowed). To compare predictive validity, Area Under Curve (AUC) statistics were calculated. Some further examinations of factors associated with reoffending were attempted.

Development of new statistical models to predict rare, harmful reoffending

Finally, statistical models were created to explore the possibility of improved predictors of certain types of rare, harmful reoffending for those with previous sanctions for these offences. These models were only fitted for arson, dangerous driving and racially aggravated offences, where the number of reoffenders was sufficiently high for a valid statistical model to be selected (see criteria in Harrell, Lee and Mark, 1996). Logistic regression models were constructed, with candidate variables including measures of static and dynamic risk both scored and unscored in OASys, and current offence characteristics.[68]

5.3 Results

PNC data processing and identification of offender groups

As set out below, dangerous driving was the most frequent (in terms of criminal history) of the six rare, harmful offence groups:

- dangerous driving (12,653 offenders; 8.8% of the sample)
- racially aggravated offending (5,377; 3.7%);
- arson (5,255; 3.6%);
- kidnapping (1,067; 0.7%);
- child neglect (840; 0.6%); and
- blackmail (582; 0.4%).

Considering the six groups together, 21,825 offenders (15.3% of the entire sample) appeared in one group, 1,792 (1.3%) appeared in two groups, 97 (0.1%) appeared in three groups and three offenders appeared in four groups. Log linear model results indicated some tendencies for offenders to be involved in multiple harmful offence types.[69] The strongest tendency was found with blackmail, where 194 (33%) of the 582 offenders had been sanctioned for at least one of the other five offence types; on the basis of statistical independence, only 96 (17%) would be expected. Kidnapping featured 301 (28%) of 1,067 with other offences types, where 174 (16%) would be expected; for arson, the equivalent figures were 1,127 (21%) against 723 (14%) of 5,255. Dangerous driving (11% actual, 9% expected) and racially aggravated offending (15% actual, 14% expected) showed weaker tendencies. Child neglect offenders were infrequently involved in other rare, harmful offences: 749 (89%) of these 840 offenders were involved in none of the other five offence types, compared with an expected 707 (84%).

Considering particular combinations of offences, dangerous driving, blackmail and kidnap appeared to be associated, especially blackmail and kidnap: this pair of offence types (sometimes with others) appeared in the criminal histories of 49 offenders, where only four would be expected on the basis of statistical independence. Examination of sanction-level data revealed 36 convictions or cautions

---

68 Forward selection at *p*=.1 was used. Static measures included the number of previous offences for each offence group, age, gender, being a first-time (proven) offender, and whether the index conviction involved the offence of interest. Dynamic measures included the eight OASys dynamic risk factor scores, and unscored items including domestic violence perpetration and victimhood, individual questions from the emotional wellbeing section, impulsive behaviour and discriminatory attitudes. Where the index conviction involved the offence of interest, it was possible to consider acknowledgement of the impact of offending, co-offender involvement, peer group influence, accepting responsibility for the current offence and whether the current offence was part of an established pattern of similar offending.

69 Note that the statistical independence condition is not perfectly upheld in this analysis, as some offenders are represented more than once in the dataset. Deviations from what might be expected on the basis of independence are therefore likely to be exaggerated compared with a one-record-per-offender approach.

involving blackmail and kidnap together, suggesting that this pair of offences occurred as part of a single criminal act.

Risk/needs profiles and criminal histories

Table 5.1 compares the demographic characteristics of offenders in each of the six offence groups. Dangerous drivers were the most exclusively male group (98% male), and among the youngest (mean age 29.8 years), as expected from a previous research finding that 75% of dangerous drivers were males aged under 30 (Department for Transport, Local Government and the Regions, 2002). They had similar mean numbers of previous sanctions to arson and blackmail offenders (15.3, 14.5 and 15.5 sanctions respectively). Child neglect offenders had relatively few previous sanctions (mean of 9.2), and half were female, compared with around one-eighth of all offenders and smaller proportions within many of the other rare, harmful offender groups. The arson offender group was comprised almost entirely (88%) of White offenders, where their ethnicity was known, while the blackmail and kidnap groups included more Black, Asian and Minority Ethnic (BME) offenders than the overall OASys sample.

The mean age of racially aggravated offenders was low (28.4 years). This may reflect the historical background of the offences concerned, which only existed in the legal sense upon implementation of the Crime and Disorder Act 1998, and also the concerns cited by Burney and Rose (2002) that offences with racial elements often lead to convictions for the ‘basic’ offence (i.e. without the racial element). A high proportion of identified racially aggravated offenders were therefore included on the basis of the additional OASys details regarding their current offences, with the extensive offending histories of some older offenders thus being effectively ignored.

Static likelihood of reoffending, as measured through the OGRS3 two-year percentage, was greatest for the dangerous driver group (mean of 68%). Child neglect offenders were the only one of the six groups to have a lower mean static likelihood of reoffending than those with no history of rare, harmful offending (47% compared to 51%).

Table 5.1 also allows comparison with different types of offenders with no history of rare, harmful offending. This suggests that arson offenders had similar OGRS scores, gender and ethnicity to other criminal damage offenders, but were older and had more extensive criminal history. Blackmail offenders had similar ethnicity and gender profiles to robbers (i.e. often BME and usually male), but were older, with more previous sanctions and higher OGRS scores. The child neglect group was confirmed as including a higher proportion of female offenders than any other group. Dangerous drivers were of similar age to other motoring offenders (not drunk drivers), but were more usually White and had more previous sanctions and thus higher OGRS scores. The kidnap and racially motivated groups had unusual profiles, though the kidnap group were like robbers in frequently being BME and having similar OGRS scores.

Table 5.2 displays the criminogenic need profiles of each group. Levels of accommodation and employability need were generally above average, though less so for the dangerous driver and racially aggravated offender groups. The child neglect group had the highest levels of relationships need (85%) – unsurprisingly, given that one of the three scored relationships questions involves current relationship with close family, which can include children – but were the only group not to have elevated levels of lifestyle/associates and attitudes needs (both 50%). Substance misuse needs varied considerably, without being extremely high or low for any group: blackmail, dangerous driving and kidnapping offender groups had relatively high levels of drug misuse need, racially aggravated offenders had more alcohol misuse need, the arsonist group had raised levels of both needs and the child neglect group had lower levels of both needs. Thinking and behaviour need levels were above average for all offender groups. RoSH ratings were greatest for those convicted of kidnapping (32% high/very high) and only slightly above the population average for those convicted of dangerous driving (8% high/very high).

Compared with conventional offence groups, blackmail offenders seemed very similar to robbers; kidnap offenders were rather less so, having more frequent relationship need and less frequent lifestyle and drug needs. Arsonists were not especially similar to other criminal damage offenders, misusing alcohol less often but having more of every other need and rated as more harmful. The profile of child neglect offenders was essentially unique. Dangerous drivers had higher levels of all ten needs and were rated as more risky than other motoring offenders, and the same is true of those convicted of racially aggravated offences – typically involving interpersonal violence – compared with other violent offenders.

Table 5.1: Demographic characteristics and static risk factors of those with a known history of each offence type

Group *(n)*	**OGRS3 2 yr %**	**Previous sanctions**	**Age**	**Gender**	**Ethnicity: % in each group**					
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**% female**	**Asian**	**Black**	**Mixed**	**Other**	**White**	**Missing / not stated**
Arson *(5,255)*	65 (20)	14.5 (10.2)	30.2 (9.9)	7	1	1	1	<1	88	8
Blackmail *(582)*	62 (21)	15.5 (9.7)	33.8 (9.2)	7	5	11	4	<1	69	11
Child neglect *(840)*	47 (24)	9.2 (9.6)	32.9 (8.7)	50	1	5	1	<1	84	9
Dangerous driving *(12,653)*	68 (20)	15.3 (10.0)	29.8 (8.8)	2	3	3	2	<1	84	7
Kidnapping *(1,067)*	57 (22)	13.3 (9.4)	32.9 (9.2)	7	8	10	2	<1	70	10
Racially aggravated offence *(5,377)*	61 (23)	11.6 (10.8)	28.4 (9.7)	13	4	3	2	1	81	9
No known history of any rare, harmful offences *(120,450)*	51 (25)	8.2 (7.9)	30.4 (10.3)	14	3	4	2	1	80	11
Of which, current offence:										
Violence against the person (32,742)	45 (22)	6.8 (6.3)	29.9 (10.0)	11	3	3	2	<1	79	13
Sexual offences (2,644)	14 (16)	3.5 (5.1)	43.5 (14.6)	2	2	1	1	1	84	11
Burglary (8,633)	67 (17)	12.9 (8.8)	27.1 (7.4)	6	1	3	2	<1	86	8
Robbery (3,398)	55 (22)	10.5 (8.1)	26.3 (7.2)	9	4	12	4	1	72	7
Theft and handling (19,870)	68 (21)	11.8 (10.1)	28.6 (8.7)	25	2	3	2	<1	84	8
Fraud and forgery (4,054)	40 (25)	5.9 (6.6)	33.6 (10.8)	36	6	7	1	1	72	12
Criminal damage (3,167)	61 (19)	8.8 (6.8)	26.7 (8.6)	7	1	2	1	<1	85	10
Drug offences (9,017)	46 (23)	8.6 (7.4)	30.4 (9.0)	16	5	7	2	1	75	10
Drink driving (12,498)	33 (21)	5.0 (5.3)	35.8 (11.1)	14	3	3	1	1	82	10
Other motoring offences (9,909)	58 (19)	8.6 (7.1)	29.1 (9.6)	7	4	6	2	1	78	9
All other offences (14,344)	46 (25)	7.1 (7.8)	31.0 (11.4)	15	3	3	2	1	79	12
All offenders (144,229)	***53 (25)***	***9.2 (8.6)***	***30.3 (10.2)***	***13***	***3***	***4***	***2***	***1***	***81***	***10***

Table 5.2: Criminogenic need profiles of those with a known history of each offence type

Group *(n)*	**% with each criminogenic need**								**% risk of serious harm**		
	Accom.	**Employ.**	**Rel.**	**Lifestyle & associates**	**Drug misuse**	**Alcohol misuse**	**Thinking & behaviour**	**Attitudes**	**Low**	**Medium**	**High / V. High**
Arson *(5,255)*	46	74	69	70	46	42	68	64	32	53	15
Blackmail *(582)*	43	75	65	71	53	30	70	67	25	57	18
Child neglect *(840)*	46	73	85	50	34	30	67	50	20	59	21
Dangerous driving *(12,653)*	37	67	54	71	47	28	63	64	45	47	8
Kidnapping *(1,067)*	47	70	72	66	45	26	68	67	14	54	32
Racially aggravated offence *(5,377)*	40	65	60	66	39	52	66	65	27	61	12
No known history of any rare, harmful offences *(120,450)*	34	55	51	52	36	32	50	45	54	40	6
Of which, current offence:											
Violence against the person (32,742)	*33*	*47*	*55*	*38*	*23*	*43*	*51*	*40*	*27*	*63*	*10*
Sexual offences (2,644)	*40*	*37*	*68*	*50*	*7*	*11*	*60*	*48*	*10*	*51*	*39*
Burglary (8,633)	*50*	*82*	*61*	*79*	*67*	*26*	*63*	*65*	*61*	*36*	*3*
Robbery (3,398)	*48*	*80*	*59*	*80*	*63*	*26*	*60*	*59*	*16*	*65*	*19*
Theft and handling (19,870)	*43*	*75*	*57*	*68*	*60*	*23*	*55*	*56*	*75*	*24*	*1*
Fraud and forgery (4,054)	*24*	*49*	*44*	*41*	*23*	*11*	*41*	*35*	*87*	*12*	*1*
Criminal damage (3,167)	*40*	*62*	*59*	*55*	*34*	*60*	*60*	*51*	*46*	*49*	*4*
Drug offences (9,017)	*33*	*63*	*44*	*67*	*73*	*11*	*42*	*45*	*77*	*22*	*1*
Drink driving (12,498)	*16*	*26*	*34*	*34*	*9*	*54*	*33*	*23*	*73*	*26*	*1*
Other motoring offences (9,909)	*24*	*51*	*41*	*51*	*26*	*21*	*51*	*47*	*74*	*25*	*1*
All other offences (14,344)	*30*	*49*	*49*	*45*	*23*	*34*	*51*	*43*	*48*	*43*	*9*
All offenders (144,229)	***35***	***57***	***52***	***54***	***37***	***33***	***53***	***48***	***51***	***42***	***7***

Note. Accom. = accommodation; Employ. = employability. Rel. = relationships. Risk of serious harm = highest community risk.

Examination of the criminal histories of the six groups shows that dangerous driving is often repeated: the average dangerous driving offender had 1.28 such sanctions, compared with 1.12 arson sanctions for those with any arson sanction, 1.09 for racially aggravated offending and 1.03 to 1.04 for the other three groups. In the most notable crossover between groups, the average blackmail offender had 0.09 kidnapping sanctions, compared with 0.01 among the whole sample. Considering histories of the other 20 offence groups, arsonists had sanction counts at least 50% higher than the whole sample for 14 groups, including over double higher for criminal damage, motor theft and burglary offences. Blackmail offenders had over double the whole-sample rates of homicide/wounding, indictable assault, robbery, threats/harassment, contact sexual offences, burglary and fraud/forgery, suggesting tendencies to commit both seriously harmful and the more lucrative offences. Dangerous drivers were above the whole-sample average for all but sexual offences, had over three times the whole-sample average for motor theft and other motoring offences and over twice the average for burglary. Kidnapping offenders had over three times the whole-sample average for homicide/wounding, robbery and contact sexual offences, and over double for threats/harassment and motor theft, suggesting a wide range of motivations. Their mean child abduction sanction count was only 0.01. Racially aggravated offenders had some violence propensity, including double the whole-sample averages for weapon possession and public order, while child neglect offenders were near or below the whole-sample averages for all 20 offence groups.

Prediction of rare, harmful reoffending using existing scales

Table 5.3 compares two-year proven reoffending outcomes for each of the six rare, harmful offences, separating those with and without any previous sanctions for the relevant offence. Reoffending rates for those with relevant previous sanctions ranged from 0.3% (blackmail) to 4.4% (racially aggravated offences), and for those without relevant previous sanctions ranged from 0.04% (blackmail) to 1.1% (racially aggravated offences). The odds ratios show strong evidence for specialisation: even after accounting for static risk, those with previous history generally had between 2.7 and 4.7 times the likelihood of new proven offending than those without previous history. Blackmail offences had even stronger evidence of specialisation (odds ratio 8.1, though with broad confidence intervals). This clear evidence of specialisation does not make it easy to isolate those likely to commit rare, harmful offences in the future, because few offenders have histories of these offences – even for dangerous driving, those with no known history of the offence still comprised almost two-thirds of the reoffenders. However, it is plausible that offenders with particular characteristics (e.g. high scores on existing scales) might be many times more likely to reoffend than other offenders.

Child neglect is an offence which is clearly influenced by opportunity: that is, sustained responsibility for the welfare of a child. In OASys terms, those with the greatest such responsibility are likely to be women who live with children (identified using OASys question 3.2). The analyses reported below on this offence group differentiate by gender and question 3.2.

Table 5.3: Proportions with proven reoffending for each offence within a two-year follow-up

Offence type	Those with no known history for this offence type			Those with known history for this offence type			Odds ratio (95% confidence interval): those with known history vs. those without	
	Number of offenders	***Mean OGRS3 score***	**% (n) proven reoffending**	***Number of offenders***	***Mean OGRS3 score***	**% (n) proven reoffending**	**Simple**	**Controlling for OGRS3 %**
Arson	*138,974*	*52*	0.21 (294)	*5,255*	*65*	0.72 (38)	3.44 (2.45, 4.82)	2.75 (1.95, 3.87)
Blackmail	*143,647*	*53*	0.04 (51)	*582*	*61*	0.34 (2)	9.71 (2.36, 40.0)	8.12 (1.97, 33.5)
Child neglect	*143,389*	*53*	0.11 (154)	*840*	*47*	0.48 (4)	4.45 (1.65, 12.0)	4.67 (1.73, 12.6)
Dangerous driving	*131,576*	*51*	0.74 (978)	*12,653*	*67*	4.17 (528)	5.82 (5.22, 6.48)	3.80 (3.40, 4.24)
Kidnapping	*143,162*	*53*	0.10 (147)	*1,067*	*57*	0.47 (5)	4.55 (1.87, 11.1)	4.31 (1.76, 10.5)
Racially aggravated	*138,852*	*52*	1.12 (1562)	*5,377*	*61*	4.43 (238)	4.07 (3.54, 4.68)	3.36 (2.92, 3.87)

Offenders with no previous history of the rare, harmful offence

Table 5.4 compares each risk predictor's AUCs for each type of rare, harmful reoffence, among offenders with no history of the relevant offence. Arson recidivism was best predicted by the OVP1 total score (AUC = 0.71), within which the dynamic subscale was also predictive (AUC = 0.70), both representing upper-moderate predictive validity. The OVP1 total score is only slightly more predictive than the dynamic subscale because static factors, especially gender, are less predictive for arson than for other violent reoffences. Kidnapping and racially aggravated offences appeared to function well as OVP1-type offences for those with no such history. Blackmail was predicted well by OGRS3 but not especially well by either OGP1 or OVP1.

Of the 8,010 women living with children but with no child neglect history, 48 (0.6%) had a child neglect reoffence. This was not significantly different from the 1.1% (2/176) rate among such women who did have a known child neglect history. It compared with 0.08% of all male or female offenders with no child neglect history,[70] and a 0.3% (2/664) rate for all other male or female offenders with child neglect history. OGRS3, OGP1 and OVP1 all had poor overall predictive validity for child neglect, as these predictors all give women slightly lower reoffending probabilities than men and do not take account of access to children. Among women who lived with children only, the OGP1 dynamic scale had the highest AUC, a moderate .64. The offender's age was also weakly predictive (AUC=.60), and a combination of age and the OGP1 dynamic score seemed most predictive (AUC=.66). For example, women aged under 25 who lived with children and scored at least 2 (of 12) points on the drug misuse scale had a 1.1% (6/526) rate. In practice, therefore, young women who live with children and have high OGP1 dynamic scores seem most at risk of committing future child neglect offences.

Static factors were most predictive of dangerous driving reoffending, with OGRS3 (AUC = 0.73) and the OVP1 static scale (AUC = 0.72) proving good predictors. Dangerous driving reoffending was extremely strongly associated with age. Among those with no such prior offending, dangerous driving reoffending rates were 1.7% for those aged 18–19, 1.3% for ages 20–21, 0.9% for ages 22–23, then declining to 0.1% for ages 46–50 and 51+. The odds ratio of 22 separating the 18–19 and 51+ age groups compared with 6.8 for OGP1-type (i.e. non-violent) and 8.6 for OVP1-type (i.e. non-sexual violent) reoffending. Dangerous driving reoffending was also extremely male-centred, with odds 7.7 times greater for men than women, compared with 1.2 for OGP1-type and 1.8 for OVP1-type reoffending. In all, 47% of dangerous driving reoffences by those without known history of this offence were committed by males aged 18–21 (who comprised 23% of the sample), 7% by males aged 36 and over (28% of the sample) and 2% by women (14% of the sample).

[70] Women living with children comprised 6% of the sample but 31% of the child neglect reoffenders.

Table 5.4: Predictive validity of existing risk assessment tools for those with no known history of each offence

Predictor (range)	**AUC (95% CI) for offence group**					
	Arson (n=138,974, 0.21% proven reoffending)	**Blackmail (n=143,647, 0.04% proven reoffending)**	**Child neglect (n=143,389, 0.11% proven reoffending)**	**Danger. driving (n=131,576, 0.74% proven reoffending)**	**Kidnapping (n=143,162, 0.10% proven reoffending)**	**Racially aggravated (n=138,852, 1.12% reoff.)**
OGRS3 2-year score (0–100)	.65 (.62, .68)	.70 (.64, .77)	.55 (.51, .59)	.73 (.71, .74)	.66 (.62, .69)	.69 (.68, .71)
OGP1 dynamic score (0–40)	.63 (.60, .66)	.58 (.50, .65)**	.59 (.55, .63)+	.61 (.59, .62)***	.62 (.58, .66)	.63 (.62, .65)***
OGP1 total score (0–100)	.65 (.63, .68)	.67 (.61, .73)*	.57 (.53, .61)++	.70 (.69, .71)***	.66 (.63, .70)	.69 (.68, .70)
OVP1 static score (0–60)	.66 (.63, .70)	.69 (.63, .75)	.51 (.46, .56)*	.72 (.70, .73)	.72 (.68, .75)+++	.71 (.70, .73)+++
OVP1 dynamic score (0–40)	.70 (.67, .73)++	.53 (.45, .61)***	.56 (.51, .60)	.54 (.52, .56)***	.63 (.59, .67)	.69 (.68, .70)
OVP1 total score (0–100)	.71 (.68, .74)+++	.64 (.58, .70)*	.53 (.48, .58)	.66 (.65, .68)***	.71 (.67, .74)++	.74 (.73, .75)+++

Note. The OGP1 static score = 0.6*OGRS3 2-year score (rounded, 0–60 range). Its AUCs are therefore very similar to those of the OGRS3 2-year score. The AUC of the OGRS3 2-year score is compared with the AUC of each other tool. Where the OGRS3 AUC is higher, *: $p<.05$. **: $p<.01$. ***: $p<.001$. Where the OGRS3 AUC is lower, +: $p<.05$. ++: $p<.01$. +++: $p<.001$.

Offenders with previous history of the rare, harmful offence

Reliable AUC estimates could only be obtained for arson, dangerous driving and racially aggravated reoffending, as there were fewer than ten recidivists for each of the other three offence types. These are presented in Table 5.5.

The arson recidivism rate was 0.7%: 38 of 5,255. The OVP1 dynamic score, with a good AUC of 0.73, was the best predictor of arson reoffending, followed by the OVP1 total score (AUC = 0.71), with the static score faring less well. Contrary to OVP1 scoring, women (1.1%; 4/371) and the oldest offenders (1.3%, 5/389, at ages 46+) had high rates. While there was only one (0.2%) reoffender among 637 offenders with no history of OVP1-type offending, the OVP1 sanction count was otherwise a weak predictor. Using the banded OVP1 score, as might be recommended in practice, those in the Low band had a 0.1% (2/1,764) rate; those in the medium band a 0.8% rate (22/2,682); those in the high band a 1.6% (11/703) rate, and those in the very high band a 2.8% (3/106) rate.

The dangerous driving recidivism rate was 4.2%: 528 of 12,563. The best predictors were the OGRS3 and OGP1 total scores, as was the case among offenders with no dangerous driving history. However, among those with history, even these predictors performed only moderately (AUCs of 0.65 and 0.64 respectively). Age and gender did have predictive power. Offenders aged 18–19 and 20–21 had 6.7% and 7.5% rates respectively, compared with under 2% for 41–45 and 46–50 and 0.3% (1/309) for those aged 51+. Rates were 0/257 for females and 1/486 for first-time offenders. Rates across the four OGP1 bands were: low, 0.7% (16/2,396); medium, 3.8% (208/5,446); high, 6.4% (254/3,956), and very high, 5.9% (50/855).

The racially aggravated offending recidivism rate was 4.4%: 238 of 5,377. The full OVP1 score was the best predictor: its AUC of 0.68 was moderate, although slightly constrained by the nature of the sample: as all criminal charges involving racial aggravation are for violence-related offences, it is unsurprising that only 4% of the sample (with a 2.5% rate: 6/239) had no OVP1 sanctions.[71] Yet the predictive validity of OVP1 for this offence rests largely on its criminal history elements: females and males alike had 4.4% rates, while there was an unusual age curve – rates fell modestly, from 4.8% at age 18–19 to 3.3% at age 31–35, then rose to over 8% for ages 46–50 and 51+. Violent criminal history was predictive: those with 0–2 OVP1-type sanctions had a 2.2% (42/1,869) rate, those with 3–6 sanctions had a 4.0% (91/2,269) rate, those with 7–10 sanctions had a 6.6% (56/852) rate and those with 11+ sanctions had a 12.7% (49/387) rate. Rates across the four OVP1 bands were: low, 2.0% (29/1,480); medium, 3.7% (99/2,668); high, 8.3% (87/1,053), and very high, 13.1% (23/176).

[71] As mentioned earlier, OASys data could be used to identify an index offence as racially motivated for the purposes of defining the sample. However, such offences are not counted as violent for OVP1.

Table 5.5: Predictive validity of existing risk assessment tools for those with known history of each offence

Predictor (range)	AUC (95% CI) for offence group		
	Arson (n=5,255, 0.72% proven reoffending)	**Dangerous driving (n=12,653, 4.17% proven reoffending)**	**Racially aggravated offences (n=5,377, 4.43% proven reoffending)**
OGRS3 2-year score (0–100)	.55 (.46, .63)	.65 (.63, .67)	.64 (.61, .68)
OGP1 dynamic score (0–40)	.54 (.46, .61)	.59 (.57, .62)***	.64 (.60, .67)
OGP1 total score (0–100)	.54 (.46, .62)	.64 (.63, .66)	.66 (.63, .69)+
OVP1 static score (0–60)	.60 (.51, .69)	.61 (.59, .64)**	.62 (.59, .66)
OVP1 dynamic score (0–40)	.73 (.65, .81)++	.55 (.52, .57)***	.67 (.64, .71)
OVP1 total score (0–100)	.71 (.64, .78)+++	.59 (.57, .61)***	.68 (.65, .72)+

Note. The OGP1 static score = 0.6*OGRS3 2-year score (rounded, 0–60 range). Its AUCs are therefore very similar to those of the OGRS3 2-year score. The AUC of the OGRS3 2-year score is compared with the AUC of each other tool. Where the OGRS3 AUC is greater, *: p<.05. **: p<.01. ***: p<.001. Where the OGRS3 AUC is lesser, +: p<.05. ++: p<.01. +++: p<.001. AUCs were only estimated for offence groups with >10 proven reoffenders, as shown in Table 5.3.

Development of new statistical models to predict rare, harmful reoffending

Logistic regression models were created for the three most frequent rare, harmful offences: arson, dangerous driving, and racially aggravated offences. All items listed in the results for these three offences were significant at $p<.05$ unless otherwise stated.

Offenders with no previous history of the rare, harmful offence

A logistic regression model was developed to predict arson reoffending among the 138,974 offenders without prior sanctions for this offence, building upon OVP1. Initial data analyses suggested that OVP1 underestimated the importance of alcohol in predicting future arson, and that lifestyle (section 7), emotional wellbeing (section 10) and thinking/behaviour (section 11) items were also worth modelling. The selected model included OVP1 and the following OASys questions:[72]

- 10.3 social isolation; odds ratio (OR) 1.4 for a score of 1 or 1.9 for a score of 2
- 10.5 self harm or attempted suicide; OR 1.2 for a score of 1 or 1.4 for a score of 2
- 10.7 childhood behavioural problems item; OR 1.5 if a problem existed
- 11.2 impulsivity; OR 1.5 for a score of 1 or 2.3 for a score of 2

The OVP1-type offence group was not uniformly predictive: controlling for the OVP1 score, weapon possession and (non-arson) criminal damage sanctions were more predictive than homicide/wounding or indictable assault sanctions (p=.07 for weapon possession; ORs of 1.2, 1.1, 0.7 and 0.7 per sanction, respectively). Moreover, new arsonists were seldom known contact sex offenders (OR 0.3 per sanction) and were less prolific generalist offenders (non-motor theft – the most frequent of all offence types – had an OR of 0.94 per sanction).

A logistic regression model of dangerous driving reoffending, among the 131,576 offenders without prior sanctions for this offence, revealed that young age, being male and not being a first-time offender all strongly increased risk (ORs of 20 (for age 18–19 vs. 51+), 4 and 4 respectively). Previous robbery / aggravated burglary (OR 1.2 per sanction; probably an indication of high antisociality), other motoring (OR 1.2 per sanction; e.g. driving whilst disqualified or uninsured) and indictable assault (OR 1.1 per sanction) also increased risk.[73] Employment, lifestyle/associates and thinking/behaviour needs were also associated with increased risk (ORs around 1.2 for each), while accommodation, alcohol misuse and emotional wellbeing needs were associated with decreased risk (OR 0.9, 0.6, 0.8 respectively).[74] This model achieved an AUC of 0.80, which is as high as the AUC for OGP1 as a predictor of any non-violent reoffending, and has not been exceeded in any large-scale OASys samples. In practical terms, it suggests that the driving habits of young prolific male offenders should be scrutinised carefully, especially if they demonstrate lifestyle and cognitive deficits and have a history of less directly harmful motoring offences.

[72] See Appendix D for a list of OASys scored questions.
[73] Drink driving was not associated with increased risk.
[74] Those with accommodation or emotional wellbeing needs may lack access to a motor vehicle or the confidence to drive one, respectively, while those who misuse alcohol might be convicted of drink driving rather than dangerous driving.

A model combining OVP1 with various criminal history and offence analysis items showed that racially aggravated recidivism was more frequent for those whose index offence had a stranger victim (OR 1.2), and those with histories of summary assault (OR 1.1 per sanction) and/or public order (OR 1.1 per sanction). (The predictive value of summary assault and public order offences was additional to their contribution to the OVP1 score.) Given the lack of dynamic risk factors associated with this offence, beyond those already in OVP1, combining OVP with clinical observance of relevant behaviours may be most appropriate for monitoring the risk of this offence.

Offenders with previous history of the rare, harmful offence

The opportunity to improve prediction was most limited for arson recidivism. The AUC of OVP1's dynamic factors was already good, at 0.73, and with only 38 recidivists the scope for reliable statistical modelling was limited. Nevertheless, a logistic regression model found some similar results to the model for those with no arson history: in addition to OVP1, the count of prior weapon possession sanctions was positively predictive (*p*=.02; OR 2.2 if any such offences[75]), and the count of prior non-motor theft was negatively predictive (*p*=.05; OR 0.9 per sanction). The drug offence count was also negatively predictive (*p*=.07; OR 0.7 per sanction), reinforcing the tie between arson and OVP1 (which stresses alcohol misuse) rather than OGP1 (which stresses drug misuse). To ensure model robustness, individual dynamic risk factor items were not modelled; instead, the model showed that those with emotional wellbeing need were more likely to reoffend (*p*=.02, OR 2.4) than those without.

A logistic regression model of dangerous driving recidivism among male offenders (as there were no female recidivists) confirmed age to be an important factor, and included sanction counts for dangerous driving (*p*=.004; OR 1.2 per sanction) and other motoring offences (*p*<.001; OR 1.1 per sanction), criminogenic need measures for education/training/employability (*p*=.002; OR 1.5) and lifestyle and associates (p=.002; OR 1.5), and the impulsivity item (*p*=.01; OR 1.2 per point). A count of sanctions which involved neither dangerous driving nor other motoring offences had a small but significant negative association with reoffending (*p*=.02; OR=0.98 per sanction), but being a first-time offender was strongly associated with a lower probability of recidivism (*p*=.02; OR=0.08). This suggests that a long history of general criminality does not make dangerous driving recidivism more likely, but that the presence of some criminal history distinguishes likely reoffenders and non-reoffenders. Overall, the model highlights the need to consider males aged 18–21 with employability, lifestyle and impulsivity needs and a history of dangerous driving.

A logistic regression model for racially aggravated recidivism achieved some improvement (AUC=0.71) upon OVP1, essentially by including the number of racially aggravated sanctions (p<.001; OR 1.6 per sanction) and not including age and gender. Violent (OVP1-type) sanctions were predictive (OR 1.06 per sanction), and non-violent sanctions only marginally predictive (OR 1.01 per sanction).

---

75 Because of the potential biasing effect of the three offenders with 10+ weapon sanctions all being arson recidivists, this item was recoded to a binary variable: i.e. no weapon sanctions vs. some weapon sanctions.

Dynamic risk was important: the odds ratio for each point on the 40-point OVP1 dynamic scale was 1.06, as much as an additional violent sanction.

5.4 Implications

Specialisation in criminal careers

The findings in this chapter support the notion of specialisation in particular types of offence. Even controlling for general criminality, as represented by the OGRS3 score, most types of rare, harmful recidivism are three or four times more likely for those with a history of that offence type, with blackmail offending still more specialised. Dangerous driving appears to be linked with a broader group of generally delinquent motoring offences (i.e. those not related to drink driving). Some linkages between the offence types were detected: many individuals with histories of both blackmail and kidnap offending were found, and dangerous driving was also somewhat associated with these two offence types. Arson and racially motivated offending were similarly linked with each other. However, child neglect stood alone from the other offence types.

There are overlaps between these findings and those presented in earlier research studies. Soothill, Francis and Liu (2008) considered the relationships between recidivism and previous convictions for arson, blackmail, kidnapping and threats to kill. (Threats to kill has always been among the broad class of offences considered as violent in OVP, on the basis of results in Howard (2009), and is not considered in this chapter.) They found that arson appeared to be the most specialised offence, while a modest but definite link existed between blackmail and kidnap. The current chapter's link between blackmail and kidnap but not arson was therefore consistent with these results.

Soothill, Ackerley and Francis (2004) considered the criminal careers of arsonists. In a 2000–01 sample, 43% had any previous conviction, 28% theft, 23% criminal damage, 20% violence, 18% motoring, 18% other, 16% burglary and 6% arson; this range of offences suggests that arsonists may split quite strongly between two groups: specialists and first time offenders, and versatile generalist offenders. The 20-year reconviction rate for arson, after conviction in 1980–81, was 10.7%, compared with our 0.7% two-year rate, suggesting a possible downward trend. Within arson, there was some specialisation of the specific offence of arson endangering life, an offence sufficiently scarce in the current study (8% of index arson convictions) that it was not treated separately.

A survey of convicted dangerous drivers (Department for Transport, 2004) has suggested that around one-third did not consider their offence aberrant in comparison with their usual driving behaviour, and around half admitted that their driving involved mistakes and errors, suggesting that a sub-group with persistently poor motoring behaviour may exist. Rose (2000), considering offenders convicted of serious motoring offences and a control group of other indictable offenders, also found evidence of specialisation, as 39% of reconvicted dangerous drivers had a new serious traffic offence compared with 14% of indictable non-motoring offenders.

Prediction and offender management: specific offence types

The findings in this chapter provide some support for using existing risk predictors to predict some types of rare, harmful offence, demonstrating that actuarial risk assessment instruments can be reasonably adept at predicting unexpected outcomes for which they were not specifically designed. Arson reoffending is well predicted by OVP, whether or not the offender has any record of this offence, and therefore OVP user guidance should be changed to include arson within its 'violent-type' or 'OVP-class' offence classification. This would have the additional benefit of simplicity – the current guidance (Howard, 2009) separates arson from other criminal damage offences. It would result in OVP1 predicted probabilities becoming slight underestimates of the revised OVP-class reoffending rates, but this inaccuracy would be minor, and would be offset for those with arson history by the increase in their counts of previous OVP-class sanctions. Moreover, the practical use of OVP1 scores to indicate relative risk would be unaffected. However, it should not be forgotten that arson reoffending rates are much higher among those with arson history than those without, and therefore any given high OVP1 score indicates greater potential to commit future serious harm among past/current arsonists. Arson is therefore specialised in the sense that offenders persist in the offence, but generalist in as much that it can be predicted using an instrument which covers a wide range of violent behaviour. While offender profiles suggest a limited similarity with other criminal damage offenders, arsonists do seem to constitute a separate group who may place greater and different demands on offender management resources.

Weapon possession appears to be a risk factor for arson among those with no history of this offence. Given that weapon possession is also a risk factor for homicide and wounding, the most serious mainstream violence offences (Howard, 2009), this finding reinforces its value as a risk indicator for future serious harm.

Dangerous driving is well predicted by the OGRS3 score, but the distinctive nature of dynamic risk for this offence means that OGP1 does slightly worse than OGRS3 while OVP1 does not do well. Advice to assessors could state that those most at risk of dangerous driving offending are young men with histories of dangerous driving or driving while disqualified or uninsured, with employability, lifestyle and impulsivity problems. The rates of dangerous driving among those with many or all of these risk factors are sufficiently high – for some individuals, above 10% in two years – that dealing with the risk of this offence could justifiably be considered a priority.

Racially aggravated reoffending is fairly well predicted by the OVP1 score, with the caveat that among those with such history, age and gender are not relevant (meaning that a moderate OVP1 score for an older and/or female offender is likely to underestimate risk) while those with multiple previous sanctions for racially aggravated offending are particularly likely to commit similar harmful reoffences. Racially aggravated offences should be part of the OVP class of offences.

The rarity of blackmail, kidnapping and child neglect offences means that conclusions about the use of risk predictors can only be drawn from analysis of those with no history of these offences. However, given that the same predictors worked for those with and without history for the more frequent arson, dangerous driving and racially motivated offence groups, it seems reasonable to extend conclusions to all offenders.

Future blackmail offences are well predicted by OGRS3. They are also well predicted by the static, but not the dynamic, parts of OVP1; those with convictions for this offence tend to have convictions for a range of violent offences. Future kidnapping is well predicted by OVP1. As with arson, kidnapping could be added to the list of OVP-class offences. While the criminal histories, demographics and risk/need profiles of both blackmail and kidnapping offenders suggest some similarities with robbers, the criminal histories of these offenders also suggest a diverse range of motivations which may require especially insightful case formulation and highly personalised sentence planning.

Among all offenders, child neglect is not predicted well by any of the predictors studied, but a combination of age and the dynamic OGP1 score predicts reasonably for women with child care responsibilities. Offender managers can therefore be advised to focus upon young women with child care responsibilities and high dynamic OGP1 scores, among whom child neglect reoffending seems frequent enough for this offence to be considered specifically. DePanfilis (2006) provides a summary of evidence on risk and protective factors for child neglect, looking at both children and parents, which may be of value to staff working with at-risk offenders. Underlying, enduring and situational factors were identified, akin to the static, stable and acute factors often discussed in violent and sexual offender risk assessment.

In summary, the implications of these findings for OGP and OVP are as follows:

- Arson, kidnapping and racially aggravated offences should be included in the OVP classification
  - no warning needs to be given about their overall predictive validity;
  - however, assessors should be aware that specialisation does exist, so those with histories of each offence are more likely to commit that offence again.
- Blackmail and dangerous driving should be included in the OGP classification, while noting the additional guidelines above regarding dangerous driving.
- Child neglect offending should be considered outside the scope of both predictors (though previous sanctions for child neglect will continue to be counted as non-violent when computing OVP scores), but see the paragraph above for important factors to be considered in identifying at-risk offenders.
- The offence classification used in the forthcoming development of version 2 of OVP (see Chapter 9) will thus include arson, kidnapping and racially aggravated offences as violent. Version 4 of OGRS (see Chapter 8) will include a separate violence predictor, which will share OVP2's classification.

- Version 2 of OGP, and OGRS4's general reoffending scale, will be developed using all reoffending without exception as their predicted outcome, but the above caveats on child neglect reoffending should be repeated in user guidance.

6. Positive, promotive and protective factors

This chapter presents the findings of a study which examined the positive, promotive and protective factors recorded within OASys. Positive factors were deemed to be ‘promotive’ when they were negatively correlated with reoffending, having controlled for risk factors. They were deemed to be ‘protective’ when moderating the impact of specific risk factors. Key points are as follows:

- The textual analysis revealed that the positive factors recorded within the OASys sentence plan correspond to the socio-economic and individual-level domains covered by the core OASys assessment. The prevalence rates of the extracted positive factor categories were relatively low, indicating that the full range of positive factors may not always have been considered.
- The optimum model for predicting reoffending included the OGRS3 score, six dynamic risk factors, five promotive factors and one protective interaction. The identification of overlapping risk and promotive processes indicates that, where risk factors are hard to change, interventions can potentially offset the risks of further offending by enhancing promotive factors, assisting with offender engagement. The interaction in the model indicated that positive family relationships moderated the impact of problematic drug misuse.
- The model combining static risk factors, dynamic risk factors, promotive factors and protective interactions performed only marginally better than a model combining static and dynamic risk factors alone. Bearing in mind that OGP1 and OVP1 have high predictive validity, this finding suggests that little would be gained, in terms of accurately predicting reoffending, from a scoring system which distinguished risk factors from promotive/protective factors.

Consideration should be given to: (i) highlighting further the importance of identifying positive as well as risk factors during OASys assessors’ training; (ii) ensuring that the recording of positive factors is carefully monitored through existing quality assurance procedures; (iii) introducing fixed response categories to encourage more systematic recording of positive factors; and (iv) distinguishing between positive factors that need to be maintained and those that need to be developed, assisting in the identification of immediately promotive/protective factors and enabling changes in status (development vs. maintenance) to be monitored.

6.1 Context

The ‘What Works’ principle of responsivity can be divided into general and specific responsivity (McGuire, 1995). While general responsivity promotes the use of cognitive social learning methods to influence behaviour, individual responsivity provides that interventions should be tailored to, amongst other things, the strengths of the offender. The consequent requirement for a structured assessment tool is to include an evaluation of personal strengths which can be integrated into the delivery of interventions. For example, within *Asset*, the assessment framework for young offenders in England

and Wales, there is a specific section on positive factors which are grouped into those relating to the individual, the family and the community (Youth Justice Board, 2003).[76] Within OASys, positive factors are considered in a less structured way through an open text field in the sentence plan which enables assessors to record ‘positive factors to be maintained or developed’.

6.2 Approach

Research questions

One of the main criticisms of the risk-needs model and actuarial risk assessment tools such as OASys is that they are overly negative, focusing upon individual’s deficiencies with insufficient attention being paid to individual’s strengths (McNeill and Weaver, 2010; Ward and Brown, 2004). To avoid seeing the worst in people and in order to maximise both engagement and responsivity to treatment, a number of commentators have argued for a shift towards a ‘strengths-based’ approach (Maruna and Le Bel, 2003) with a greater focus upon ‘desistance-related’ factors (Farrall, 2002).

This chapter examines the positive factors recorded by assessors within the OASys sentence plan. Similarly to risk factors, positive factors cover both internal assets and external strengths and they can have different degrees of dynamism, with some being more susceptible to change than others (see Chapter 4 for a change analysis). Consideration is given to how closely the positive factors recorded within the sentence plan map onto the structure of the core OASys assessment. Attention is then given to which positive factors are negatively correlated with reoffending when controlling for risk factors, thus amounting to ‘promotive’ factors, and which moderate the impact of specific risk factors, thus amounting to ‘protective’ factors. Finally, the predictive validity of these factors is examined.

Importantly, positive factors do not necessarily amount to ‘promotive’ or ‘protective’ factors. The terminology of protective factors has not always been applied consistently in the literature (Jones and Brown, 2008), with protective factors being defined “both as the absence of risk and something conceptually distinct from it” (Office of the Surgeon General, 2001). Risk and protective factors have sometimes been viewed as the opposite ends of a continuum, with the positive end of the risk dimension being negatively correlated with criminal outcomes. Alternatively, protective factors have been viewed as those factors that reduce the probability of criminal outcomes though interactions with one of more of the risk factors, reducing their influence. For consistency and clarity of understanding, a number of commentators (see Jones and Brown, 2008) have recommended that only the latter factors be termed ‘protective’, with the former definition being used to describe ‘promotive’ factors.[77]

[76] Another example is the Inventory of Offender Risk, Needs, and Strengths (IORNS; Miller, 2006) which has a Protective Strength Index, encompassing 26 items across two sub-scales: (i) personal resources (19 items) and (ii) environmental resources (seven items). The latter focuses upon instrumental and emotional support from family and friends, while the former is broken down into the following three sub-scales: cognitive/behavioural recognition (nine items assessing the ability to regulate feelings, cognitions and behaviours), anger regulation (five items reflecting the ability to regulate anger and temper) and education/training (five items reflecting the obtained level of education and training for employment purposes). As scores increase on the scales, the level of protection against reoffending is judged to increase.

[77] ‘Promotive’ factors have also been termed ‘compensatory’ factors (Luthar, 1993).

In essence, the distinction depends upon whether the positive factor has a direct effect on the outcome variable or moderates the effects of the risk variables via an interactive relationship – the former being seen as promotive and the latter protective (Schoon, 2006). Promotive factors can thus be combined into cumulative main effects regression models, with the combined promotive factors reducing the probability of criminal outcomes for those exposed and those not exposed to various risk factors. In interaction effects regression models, in contrast, the beneficial effects of protective factors are restricted to those individuals with specific risk factors.

To summarise, the four key research questions addressed in this chapter are as follows:

1. What are the main positive factors recorded within the OASys sentence plan?
2. Which positive factors are promotive, negatively correlated with reoffending when controlling for risk factors?
3. Which positive factors are protective, reducing the likelihood of reoffending by moderating the impact of specific risk factors?
4. How well do the protective and promotive factors add to the prediction of reoffending?

The samples

To examine the positive factors recorded within the OASys sentence plan, assessments completed by the probation service during 2008 were extracted from the O-DEAT database.[78] These assessments were filtered to ensure that the following standards of data completion had been satisfied:

- Each of the scored sections (1 to 12) within the core OASys assessment must have had at least four-fifths of their scored items completed – ensuring that each criminogenic need was assessed properly.
- In the Risk of Serious Harm (RoSH) component of OASys, the screening must have been completed, the decision whether to complete a full risk analysis must have been consistent with the information provided, and the four ratings of RoSH in the community must have been completed.
- In the sentence plan, a criminogenic need must have been recorded within the 'objectives and plans' section, and text of at least ten characters must have been entered in the positive factors field.

The sample was further restricted to the earliest valid assessment for each offender. This sampling left 132,093 assessments from all 35 probation trusts.

[78] While the O-DEAT database also includes assessments completed by the prison service, the vast majority of assessments are completed by probation assessors and the predominance of such assessments has increased with the rollout of Offender Management and its requirement for assessments to be completed by community-based Offender Managers.

An older sample was used to examine the associations with reoffending. The initial sampling frame was extracted from the O-DEAT database, selecting assessments completed between April 2002 and February 2006 inclusive. These assessments were filtered to ensure that a sentence date and all scored items had been recorded and that the RoSH component had been completed as set out for the sample above. The assessments were then de-duplicated, leaving one assessment per offender per sentence, and matched with records in the Police National Computer (PNC) research database.[79] It was checked that: (i) the cases could be followed up for 24 months at liberty from the date of the community sentence or discharge from custody, allowing three months for sentence and data entry to occur; and (ii) the confirmed community sentence/custodial release date and the OASys completion date were within 90 days of each other (with a further check to ensure that the nearest assessment to each community sentence/custodial release date was selected). This left a final sample size of 91,464 cases for 83,524 different offenders (representing 23% of the offenders in the initial sampling frame) for use in the analysis.

Analysis

To process the textual data recorded within the positive factors field of the OASys sentence plan, a linguistic-based text mining tool was used,[80] employing advanced linguistic technologies and natural language processing. Key concepts/terms, representing the essential information, were extracted automatically, with normalisation and grouping techniques correcting punctuation and spelling errors respectively. Closely related concepts were then grouped into higher-level categories, firstly through further linguistic based methods, identifying synonym and hyponym relationships and root terms, and then manually. The prevalence rates of the extracted categories across offender sub-groups were compared, and the categories matched against the individual-level and socio-economic domains covered by the core OASys assessment.

The ten individual-level and socio-economic sections within the core assessment, have scaled scores. Previous analysis has focused upon the associations between these ten scaled scores and proven reoffending (Howard, 2006; Moore, 2009b). However, as recognised by Stouthamer-Loeber *et al*. (2002:112), "Such analyses do not indicate whether the association is linear or lodged on either end. Further, a regression strategy with continuous variables does not clarify strengths and weaknesses of individuals." Consequently, these authors suggested an alternative approach which was "to examine the effects of protective and risk factors as represented by opposite poles of the same variable, allowing a variable to have a risk effect for one participant and a protective effect for another, depending on whether a participant scores closer to one or the other pole on the variable."

---

79 PNC numbers were recorded within OASys for most offenders, and an automatic matching procedure found reliable PNC numbers for most of the remaining cases. Cases in which the PNC did not record the offender's sex or recorded an unfeasible date of first or current conviction were rejected.

80 The text mining tool was a component of IBM SPSS Modeller.

In a number of papers (e.g. Stouthamer-Loeber, 1993; Farrington *et al*., 2008), this has been achieved through trichotomisation of the scaled variables, enabling comparisons between (a) those with high and medium scores and (b) those with low and medium scores. Adhering to this approach, the ten OASys scales were trichotomised, with scores of 0 representing potentially promotive/protective factors and scores of at least 4 for the shorter scales and 7 for the longer scales representing potential risk factors.[81] In other words, distinctions were made between offenders with no problems, those with some problems and those with significant problems within each domain.[82] The questions included within each domain (see Appendix A) were those which have previously been found to (i) contribute to each scale's internal reliability and (ii) maintain construct validity by loading onto corresponding factors (Moore, 2009b).

Chi-square tests were initially used to assess which sections were significantly associated with reoffending at both the risk and promotive ends of their scales, comparing (a) those offenders with significant problems and those with some problems and (b) those offenders with no problems and those with some problems. To account for the relationships between the factors, logistic regression models were then used. Dichotomous promotive and risk variables were created, with the promotive variables distinguishing those offenders with no problems from all other offenders (those with some or significant problems) and the risk variables distinguishing those offenders with significant problems from all other offenders (those with no or some problems).[83]

The full sample was divided into construction (60%; n=54,980) and validation samples (40%; n=36,484), with the modelling conducted on the construction sample. In all the models set out in this chapter, the independent variables were entered using a forward stepwise approach, incorporating the most significant variables in turn and then removing them at a later stage if necessary.[84] The first model focused solely upon positive factors, assessing which remained negatively correlated with reoffending when taking into account the relationships between them. Further models entered the positive factors alongside the static and dynamic risk factors, with the static factors being represented by the OGRS3 score which uses criminal history and offender demographic data to provide a percentage prediction of proven reoffending within two years. A main effects model examined which factors had promotive effects for those with or without various risk factors, while a combined main effects and interactions model established which factors interacted with the risk factors, thus amounting to protective factors for those individuals with specific risks.

81 The cut-off points for identifying risk were set at 4+ for the six point scales, 5+ for the eight point scales, 6+ for the ten point scales and 7+ for the twelve point scales, i.e. as close as possible to the 60th percentile of each scale.

82 In a number of research papers (e.g. Stouthamer-Loeber *et al*., 2002; Farrington *et al*., 2008), the variables were trichotomised as closely as possible to the 25th and 75th percentiles of the distribution. Such an approach was not possible for the OASys section scores, with over half of the offenders scoring zero on some scales.

83 To assess whether there was any problem of mulitcollinearity, the correlations between the dichotomous promotive and risk factor variables were checked through tolerance and variation inflation factor (VIF) statistics.

84 The forward stepwise approach was considered appropriate as the analysis was exploratory in nature and there was no definitive research evidence as to the relative theoretical importance of the various independent variables.

To assess whether the inclusion of promotive and protective factors improved the prediction of proven reoffending, further logistic regression models were created which were confined to the static and dynamic risk factors. In total, six logistic regression models were compared:

- static risk factors (OGRS3) model;
- dynamic risk factors model;
- positive factors model;
- combined static risk and dynamic risk factors model;
- combined static risk, dynamic risk and promotive factors model; and
- combined static risk, dynamic risk, promotive and protective factors model.

Predicted reoffending rates were calculated from each model and their accuracy checked using the validation sample. Area Under Curve (AUC) statistics were used to check that higher predicted scores represented a higher likelihood of reoffending. A weakness of AUC statistics is that they derive from the relative rankings of offenders – if one added 20 per cent to every offender's prediction, the AUC for the sample would not change, even though the proven reoffending rate would be severely overestimated. AUC statistics thus need to be supplemented by comparisons of actual and predicted proven reoffending rates. A further measure of accuracy was thus provided by the percentages correctly predicted (see Copas, 1992, *unpublished*). These values are calculated by dividing the predicted values into 'high' and 'low' at a point corresponding to the proportions who actually reoffend, and then treating all 'high' scores as predicting reoffending and all 'low' scores as predicting non-reoffending. High scorers who reoffend and low scorers who do not reoffend are then counted as correct predictions.

Limitations

The results of the linguistic text mining, used to extract the textual information recorded within the positive factors field of the sentence plan, are dependant upon the linguistic resources used. Further editing of these resources through multiple iterations could improve the accuracy and value of the concepts extracted. The distinctions that are made between promotive and protective factors are dependent upon the identification of main effects and interaction effects in the logistic regression modelling. While the inclusion of interactions in such models is an accepted way of demonstrating protective effects, strengthening of the findings through replication in other studies has been recommended (Luthar, 1993; Jessor, Turbin and Costa, 1998).

6.3 Results

Key positive factors recorded within the OASys sentence plan

Having extracted the key concepts from the positive factors text field of the 2008 assessments, these concepts were grouped into higher-level categories representing specific positive factors. The most prevalent such factors, all recorded in at least five percent of assessments, are set out in Table 6.1.[85] These factors map very closely onto the domains within the core OASys assessment which has specific sections on accommodation, education, training and employment (ETE), relationships, drug misuse, alcohol misuse and attitudes, with the latter section having a specific question on motivation (Q12.8 ‘Is the offender motivated to address offending’).

It is notable that the prevalence rates of the extracted factors were relatively low, with the most prevalent category of motivation being recorded in approximately one in five (21%) of the cases. There was also a degree of overlap between the categories, e.g. motivation being linked to drug misuse and accommodation being linked to family relationships. Much higher prevalence rates were reported from the early use of the youth justice tool *Asset*, in which practitioners are required to consider whether specified positive factors are present. This more directive approach resulted in prevalence rates of 74% for living arrangements, 74% for family/personal relationships, 59% for motivation, 54% for attitudes and thinking, and 50% for education and employment (n=3,010; Baker *et al*., 2002).

85 To assess whether the prevalence rates differed significantly between offender sub-groups, the independent grouping variables were entered into a logistic regression model (thus accounting for the relationships between the variables). Odds ratios are also presented in Table 6.1 as an indication of effect size, comparing the odds of the factor being included. In this instance, odds ratios of more than one indicate that the factor was more likely to be included for offenders within the sub-group compared to offenders within the designated reference group.

Table 6.1: Prevalence rates of main positive factor categories by offender sub-groups

	n	% with positive factor category						
		Motivation	**Employment**	**Family**	**Drugs**	**Alcohol**	**Accommodation**	**Attitude**
All	132,093	20.7%	17.2%	14.7%	11.5%	9.1%	7.4%	7.4%
Gender								
Male	115,208	20.7%^	17.9%^	14.7%^	11.3%^	9.2%^	7.4%^	7.4%^
Female	16,877	20.4%	***12.4%* (0.693)***	14.5%	***12.6%* (1.116)***	9.0%	6.8%	***6.9%* (0.930)***
Age								
18–20	22,488	19.4%^	20.8%^	15.9%^	7.2%^	8.7%^	7.8%^	7.9%^
21–24	24,058	***21.0%* (1.082)***	20.3%	16.6%	***10.1%* (1.407)***	9.1%	7.7%	8.1%
25–40	61,061	***21.2%* (1.091)***	***16.1%* (0.773)***	***14.6%* (0.916)***	***15.2%* (2.122)***	8.8%	7.4%	7.1%
41+	24,223	20.3%	***13.8%* (0.666)***	***11.9%* (0.745)***	***7.7%* (1.076)***	***10.5%* (1.211)***	***6.6%* (0.846)***	***6.7%* (0.855)***
Ethnicity[86]								
White	100,399	20.8%^	17.2%^	14.7%^	12.2%^	10.1%^	7.8%^	7.5%^
Black	8,109	20.9%	18.5%	***13.5%* (0.914)***	***11.5%* (0.948)***	***3.2%* (0.320)***	***5.6%* (0.716)***	***6.4%* (0.859)***
Asian	5,009	19.2%	18.1%	***17.7%* (1.203)***	***10.8%* (0.890)***	***5.9%* (0.584)***	***5.3%* (0.674)***	***6.2%* (0.827)***
Mixed	3,330	21.1%	17.4%	14.9%	11.9%	***5.6%* (0.554)***	7.0%	7.1%
Other	665	17.7%	15.8%	13.1%	8.7%	***5.6%* (0.552)***	5.7%	7.1%
Likelihood of reconviction[87]								
Low	35,502	19.0%^	19.6%^	15.9%^	3.4%^	8.3%^	6.7%^	8.7%^
Medium	67,365	***21.2%* (1.114)***	***17.3%* (0.883)***	***14.7%* (0.928)***	***11.5%* (3.347)***	***10.1%* (1.208)***	***7.3%* (1.083)***	***7.1%* (0.815)***
High	29,226	***21.4%* (1.128)***	***14.2%* (0.723)***	***13.1%* (0.825)***	***21.3%* (6.232)***	***7.9%* (0.951)***	***8.4%* (1.255)***	***6.2%* (0.706)***
Risk of serious harm								
Low	45,249	19.0%^	17.3%^	14.8%^	13.0%^	7.0%^	7.7%^	7.8%^
Medium	72,786	***21.3%* (1.123)***	17.3%	15.0%	***11.0%* (0.842)***	***10.5%* (1.510)***	7.7%	7.1%
High/very high	14,058	***22.9%* (1.211)***	16.9%	12.5%	***9.0%* (0.689)***	***8.9%* (1.274)***	***4.6%* (0.592)***	7.0%

^ Used as reference group within logistic regression. Asterisks indicate whether groups differ significantly (confidence level *<.001) with accompanying odds ratio vs. reference group in brackets.

[86] Ethnicity was unrecorded in over 14,000 cases (11% of the sample).
[87] The likelihood of reconviction bands are based upon the initial 0-168 OASys scoring system (replaced by OGP1 and OVP1 in August 2009).

Motivation

Having the motivation to avoid further offending has been identified as a key factor in explaining desistance (Farrall, 2004), and, as shown by Table 6.1, the most prevalent extracted category was motivation, recorded as a positive factor to be maintained or developed for approximately one in five (21%) of the offenders. The recording of this factor is likely to have been encouraged by the requirement for assessors to consider motivation in a preceding question within the OASys sentence plan and the fact that motivation is an 'interpersonal phenomenon' which is 'modifiable' (López-Viets, Walker and Miller 2002:17). It was more commonly recorded for those with a high likelihood of reconviction or for those who presented a high/very high RoSH, seemingly reflecting the importance of motivation in addressing the most entrenched offending behaviour and the need for assessors to record factors to be 'developed' as well as those to be 'maintained'. Whilst the reported motivation was sometimes a generic one of wanting to stop offending, it was in some cases more specific, relating to specific problem areas such as alcohol and drugs or a desire to fulfil parental responsibilities. Examples of entries were as follows:

- "Expressed motivation to address offending and willingness to attend additional appointments."
- "He has shown real motivation since being released, organising appointments at the job centre and trying to enrol on his college course."
- "Current level of motivation to remain in community and make positive changes in life."
- "Motivation to address her alcohol usage."
- "Motivation to abstain from class A drugs, to find employment."
- "Motivation to be better role model for child."

Employment

The category of employment, recorded in 17% of the assessments, encompassed both stable and secure employment as well as new career skills, opportunities and prospects. It was less commonly recorded for female offenders, older offenders and those with a high likelihood of reconviction. Examples of entries were as follows:

- "He has been in employment since leaving custody and is positive about remaining with company and maintaining his career with them."
- "X talked positively about his time in work and getting involved in his dad's carpet fitting business."
- "Qualified plumber and access to agencies to find work."
- "X started a three year mechanics course at Y college and appears motivated to complete… He is hoping to secure part time work with his uncle."
- "There is the Y project… It involves volunteering in the kitchen and front of house of a café for one or two days per week. Mr X seems to have a passion for cooking so if he keeps up his commitment to this project it may open doors for him into a new career that he will enjoy."

Family

The category of family, recorded in 15% of the assessments, encompassed generic family support, connections, stability and security as well as references to positive and/or improved relationships with specific family members including wives, husbands, partners, parents, siblings, sons and daughters. Positive family relationships were less commonly recorded for older offenders, Black offenders and those with a high likelihood of reconviction, and more commonly recorded for Asian offenders. Examples of entries were as follows:

- "He has a very good support system in place, in particular his mother and girlfriend who have both remained highly supportive throughout."
- "X has a good relationship with his family who are supportive and encouraging him to stay clean and away from offending."
- "Positive maintenance of family ties throughout custody."
- "Currently with his partner who is encouraging him to address his offending behaviour."
- "He is in a stable relationship…and has the responsibilities of being a father."
- "Good relationships with his younger brother… Just starting to gently build his relations back up with his father."

Drugs

The category of drugs, recorded in 11% of the assessments, covered engagement with agencies and treatment programmes, negative drug tests and current abstinence. It was less commonly recorded for Black and Asian offenders and those presenting a high/very high RoSH, and more commonly recorded for female offenders, those aged 25–40 and those with a high likelihood of reconviction. The reference to drugs in more than one in five (21%) of those cases in which the offender presented a high likelihood of reconviction is a strong indication that drugs had at some point been a problem, in many cases very recently, and that this positive factor required further development.

Alcohol

The category of alcohol, recorded in nine percent of the assessments, covered engagement with agencies, support groups and treatment programmes, current abstinence or reduced/moderate levels of consumption. It was less commonly recorded for Black, Asian and Minority Ethnic (BME) offenders and more commonly recorded for older offenders and those presenting a medium or high/very high RoSH, the latter seemingly reflecting a past link between alcohol misuse and serious harm. Examples of entries relating to drug and alcohol use were as follows:

- "Focused on rebuilding his life which does not involve drugs."
- "He states he is now clean and had no intention of returning to drug use."
- "Counselling for drugs misuse."
- "X has detoxed in prison and was drug free on release."
- "Continues to provide negative samples for Class A drugs."

- “X has reduced his alcohol intake significantly and said he has not used cocaine since the offence.”
- “She has made considerable progress in reducing her alcohol intake and she had approached Alcoholics Anonymous and her GP regarding further support.”

Accommodation

The category of accommodation, recorded in seven percent of the assessments, focused upon the existence or obtaining of stable, secure, permanent or alternative accommodation. In some instances, the reference to accommodation was linked to the existence of supportive family members, while, in other instances, it was linked to the breaking of negative peer relationships. The category was less commonly recorded for older offenders, Black and Asian offenders and those presenting a high/very high RoSH, but more commonly recorded for those with a high likelihood of reconviction. Examples of entries were as follows:

- “Accommodation has been offered through supportive family members.”
- “Finding suitable accommodation at X Guest House.”
- “Current stable accommodation with plans to move to mother’s when leaves hostel.”
- “Appears satisfied with his current accommodation – he is living with his father.”
- “Wants to find suitable accommodation to be away from previous peers and drug users.”
- “Accepted that he will have to stay in approved accommodation and not return home for a while and states that he sees this as a ‘new start’.”

Positive, responsible and/ or cooperative attitude and outlook

A positive, responsible and/or cooperative attitude and outlook was recorded in seven percent of the assessments. This positive category was less commonly recorded for female offenders, older offenders, Black and Asian offenders, and those with a high likelihood of reconviction. Examples of entries were as follows:

- “He has demonstrated a positive change in attitude and a mature pro-social outlook. He does not want to return to a life of offending.”
- “Has a positive outlook on life and his aspirations appear realistic. Is able to identify the areas of his life which he will need to address in order to avoid further offending.”
- “States that he regrets his actions and doesn’t want to get into trouble again… ‘I now have a better attitude and have matured’.”
- “He has stated that he wants to remain offence free on release and has stated his intentions to abide by any restrictions and expectations placed upon him. He has shown a willingness to engage in and complete offence focussed work and has presented as having a positive attitude towards his licence. Has been able to identify factors related to past offending and appears motivated to address these.”

Positive factors negatively correlated with reoffending (promotive factors)

The further analysis looking at the associations with reoffending used the sample of 91,464 assessments completed between April 2002 and February 2006. Approximately half (49%) of this sample were recorded as having reoffended over a two-year follow-up period. The analysis focused upon information recorded within the core OASys assessment rather than the extracted categories from the textual data in the sentence plan. The reasoning for focusing upon the core assessment was threefold:

- The extracted categories set out above mapped closely onto the individual-level and socio-economic domains covered by the core assessment.
- The prevalence rates of the extracted categories were relatively low, indicating that the full range of positive factors may not always have been considered.
- In some instances, the textual information reflected positive factors that needed to be developed rather than maintained, with the offenders having had recent problems in these areas, e.g. alcohol and drug misuse.

The analysis sought to establish whether the ten core assessment individual-level and socio-economic scales had both risk and promotive effects as reflected by scores at the two ends of the scales. The ten scales were trichotomised as set out in Table 6.2 below. As can be seen, the proportion of offenders with no problems ranged from 11% for thinking and behaviour to 60% for drug misuse, while the proportion of offenders with significant problems ranged from 6% for attitudes to 29% for thinking and behaviour.

Table 6.2: Degree of problems by OASys sections

Section	*Scale*	*% no problems*	*% some problems (Score range)*	*% significant problems (Score range)*
Accommodation	*0–8*	*53%*	*29% (1 – 4)*	*18% (5 – 8)*
ETE	*0–12*	*22%*	*52% (1 – 6)*	*26% (7 – 12)*
Financial management	*0–8*	*34%*	*49% (1 – 4)*	*17% (5 – 8)*
Relationships	*0–6*	*31%*	*49% (1 – 3)*	*20% (4 – 6)*
Lifestyle and associates	*0–6*	*29%*	*52% (1 – 3)*	*20% (4 – 6)*
Drug misuse	*0–10*	*60%*	*26% (1 – 5)*	*14% (6 – 10)*
Alcohol misuse	*0–10*	*42%*	*31% (1 – 5)*	*26% (6 – 10)*
Emotional wellbeing	*0–12*	*36%*	*51% (1 – 6)*	*13% (7 – 12)*
Thinking and behaviour	*0–12*	*11%*	*59% (1 – 6)*	*29% (7 – 12)*
Attitudes	*0–10*	*23%*	*70% (1 – 6)*	*6% (7 – 12)*

Chi-square tests were used to assess which domains were significantly associated with reoffending at both ends of their scales. Comparing those offenders with no problems to those with some problems, all ten factors were associated with reoffending. Odds ratios are also presented in Table 6.3 as an indication of effect size, comparing the odds of reoffending between the groups. As shown, the greatest divergence at this end of the scale was for ETE with an odds ratio of 0.56. Comparing those offenders with significant problems to those with some problems, all the factors were associated with reoffending except for emotional wellbeing. As shown by Table 6.3, the attitudes scale had the highest odds ratio at this end of the scale, with the odds of reoffending for those with significant problems being 1.4 times higher than those with some problems.[88]

Table 6.3: 24-month reoffending rate by level of problems across sections

Section	24-month reoffending rate for offenders		
	with no problems (Odds ratio vs. some problems)	with some problems	with significant problems (Odds ratio vs. some problems)
Accommodation	41%* (0.743)	56%	62%* (1.126)
ETE	27%* (0.557)	49%	68%* (1.381)
Financial management	37%* (0.715)	52%	64%* (1.231)
Relationships	40%* (0.770)	51%	58%* (1.135)
Lifestyle and associates	31%* (0.614)	51%	71%* (1.391)
Drug misuse	38%* (0.627)	61%	73%* (1.197)
Alcohol misuse	46%* (0.966)	48%	56%* (1.188)
Emotional wellbeing	45%* (0.871)	51%	52% (1.018)
Thinking and behaviour	29%* (0.625)	47%	62%* (1.337)
Attitudes	31%* (0.596)	53%	73%* (1.393)

Asterisks indicate whether rates differ significantly (confidence level <.001) between the extreme groups (no/significant problems) and the middle group (some problems).

To account for the relationships between the positive factors, the cases within the construction sample (n=54,980) were selected and all ten dichotomous positive factor variables (distinguishing those with no problems from all other offenders) entered into a logistic regression model. As shown by Table 6.4, eight of the positive factors were included in the model. The two excluded factors were relationships and emotional wellbeing. The odds ratios, set out in the final column of Table 6.4, are an indication of effect size, grouping the offenders by their scored positive factors and comparing the odds of reoffending between the groups.[89] In this instance, an odds ratio of less than one indicated that reoffending was less likely for those offenders with the positive factor. As can be seen, the odds of reoffending for those without any drug misuse problems were less than half the odds of reoffending for those with at least some problems.

[88] The relatively small proportions of offenders scored as having no attitudes problems and significant attitudes problems will have assisted in producing clearly divergent reoffending rates across this domain.

[89] For example, if two specific groups had reoffending rates of 40% and 60%, their corresponding odds of reoffending would be two-thirds (0.4/(1-0.4)) and 1.5 (0.6/(1-0.6)) respectively. Consequently, the odds ratio would be less than half (0.67/1.5 = 0.44).

Table 6.4: A positive factors model for predicting reoffending

Positive factor	Parameter estimate	Standard error of estimate	Odds ratio
Accommodation	-.242	.019	.785
ETE	-.593	.025	.553
Financial management	-.145	.021	.865
Lifestyle and associates	-.432	.023	.649
Drug misuse	-.715	.020	.489
Alcohol misuse	-.152	.019	.859
Thinking and behaviour	-.089	.033	.915
Attitudes	-.489	.024	.613
Constant	*1.005*	*.015*	*2.732*

Odds ratios are compared with reference categories of no identified positive factors.

To assess which of the positive factors had promotive effects for those with or without various risk factors, a main effects logistic regression model was used with the positive factors entered alongside the static and dynamic risk factors.[90] As shown by Table 6.5, the resultant model included the static risk OGRS3 score, seven dynamic risk factors and seven promotive factors. The three excluded risk factors were financial management, relationships and emotional wellbeing. The latter two domains were also excluded as promotive factors, alongside thinking and behaviour. The odds ratios, set out in the final column of Table 6.5, were greater than one for all risk factors, indicating that reoffending was more likely for those offenders with each risk factor, and less than one for all promotive factors, indicating that reoffending was less likely for those offenders with each promotive factor. Reoffending was thus linked to both the presence of risk factors and the absence of promotive factors.

Table 6.5: A static risk, dynamic risk and promotive factors model for predicting reoffending

Factor	Parameter estimate	Standard error of estimate	Odds ratio
Static risk (OGRS3)	.038	.001	1.038
Dynamic risks			
Accommodation	.147	.030	1.158
ETE	.085	.025	1.089
Lifestyle and associates	.088	.029	1.092
Drug misuse	.196	.034	1.216
Alcohol misuse	.145	.028	1.156
Thinking and behaviour	.068	.024	1.071
Attitudes	.091	.043	1.095

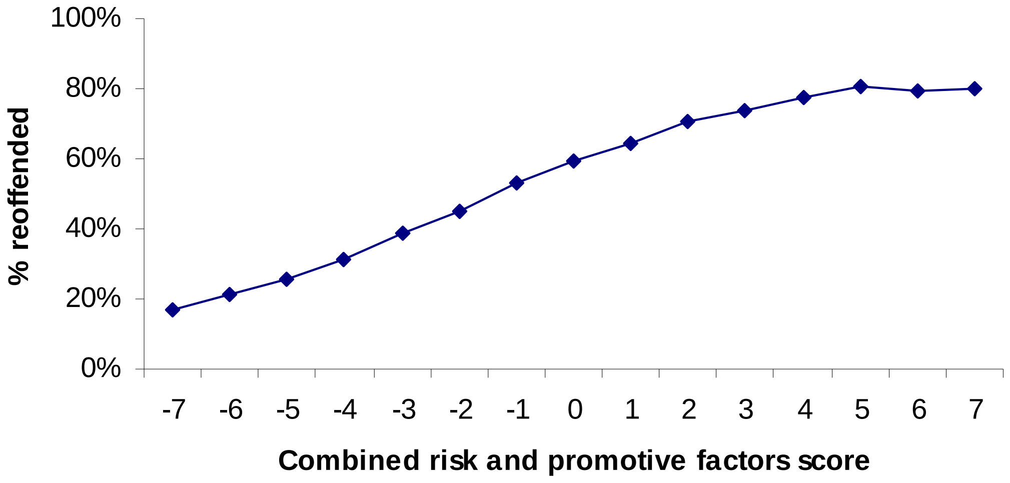
[90] The correlations between the risk and positive factor variables were checked through tolerance and variation inflation factor (VIF) statistics. The lowest tolerance value was .587 and the highest VIF value was 1.703, indicating that mutlicollinearity did not appear to be a problem.

Factor	Parameter estimate	Standard error of estimate	Odds ratio
Promotive			
Accommodation	-.135	.023	.874
ETE	-.134	.028	.875
Financial management	-.067	.023	.935
Lifestyle and associates	-.055	.025	.947
Drug misuse	-.170	.024	.843
Alcohol misuse	-.054	.025	.947
Attitudes	-.215	.026	.806
Constant	*-1.875*	*.046*	*.153*

Odds ratios are compared with reference categories of no identified risk/promotive factors and an OGRS3 percentage of 0%.

Looking at the full reoffending sample (construction and validation samples combined), only three of the 222 offenders with all seven of those risk factors included in the above model had any of the seven promotive factors included in the model. However, over two-fifths (43%; n=1,195) of those with six of the seven risk factors and three-fifths (60%; n=2,921) of those with five of the risk factors had at least one promotive factor. Clearly, therefore, promotive factors were evident for many offenders despite the presence of numerous risk factors. A combined risk and promotive factors score was created by deleting the number of promotive factors from the number of risk factors, thus producing a scaled score for each offender from -7 to 7. As shown by Figure 6.1, the reoffending rate increased across the scale from 17% for those offenders with all seven promotive factors and no risk factors (score of -7) to 80% for those with all seven risk factors and no promotive factors (score of 7). Put simply, the greater the number of risk factors and the fewer the number of promotive factors evidenced by the offender, the more likely he or she was to reoffend.

Figure 6.1: Reoffending rate by combined risk and promotive factors score



Positive factors moderating the impact of specific risk factors (protective factors)

Having established which domains had risk and promotive effects, the analysis then focused upon establishing which, if any, of the positive factors protected against specific risks. A further logistic regression model was thus created which included all possible positive and dynamic risk factor interactions.[91] Once again the OGRS3 score was entered to control for static criminal history and offender demographic factors. As shown by Table 6.6, the final model included the static OGRS3 score, six dynamic risk factors, five promotive factors and one protective interaction.

In comparison to the main effects model (see Table 6.5 above), attitudes was no longer included as a risk factor, while (i) lifestyle and associates and (ii) alcohol misuse were removed as promotive factors. Notably, relationships was added as a protective factor, moderating the impact of the drug misuse risk factor. In other words, when controlling for other static risk, dynamic risk and promotive factors, offenders with significant drug misuse problems were less likely to reoffend when they had protective relationships. This interaction indicates that the impact of drug misuse can be tackled both directly, by addressing the problem itself, and indirectly, by maximising this protective factor (Jessor, Turbin and Costa, 1998).

[91] Due to the large number of interactions entered into the model, increasing the likelihood of finding significant effects by chance, the entry criterion for the stepwise logistic regression model was amended from a score statistic of .05 to .01 and the removal criterion was amended from 0.1 to .05.

Table 6.6: A static risk, dynamic risk, promotive and protective factors model for predicting reoffending

Factor	Parameter estimate	Standard error of estimate	Odds ratio
Static risk (OGRS3)	.038	.001	1.039
Dynamic risks			
Accommodation	.146	.030	1.157
ETE	.087	.025	1.091
Lifestyle and associates	.102	.028	1.108
Drug misuse	.227	.036	1.255
Alcohol misuse	.186	.022	1.204
Thinking and behaviour	.080	.024	1.083
Promotive			
Accommodation	-.135	.023	.874
ETE	-.136	.028	.873
Financial management	-.070	.023	.932
Drug misuse	-.174	.024	.841
Attitudes	-.224	.025	.799
Protective			
Relationships promotive * drug misuse risk	-.190	.072	.827
Constant	*-1.936*	*.041*	*.144*

Odds ratios are compared with reference categories of no identified risk/promotive factors and an OGRS3 percentage of 0%.

The promotive and protective factors included in this final model support previous research claims that desistance is the result of a combination of individual choices, e.g. not engaging in drug misuse, situational contexts, e.g. positive support, and structural influences, e.g. employment opportunities (Laub and Sampson, 2001). The inclusion of attitudes as a promotive factor also points to the importance of offenders maintaining a positive frame of mind (LeBel *et al*., 2008).

The predictive validity of the protective and promotive factors

The accuracy of the logistic regression models was checked using the validation sample. As shown by Table 6.7, the final regression model, combining promotive and protective factors with static risk and dynamic risk factors, achieved a high level of discrimination with an AUC score of 0.783.[92] In other words, nearly eight out of ten randomly selected reoffenders had higher scores than randomly selected non-reoffenders. Comparing this model with one combining static and dynamic risk factors, this AUC score represented a very small improvement of .002. The model based purely upon positive factors achieved a higher level of discrimination than a model based purely upon dynamic risk factors, indicating that the absence of problems across domains was more predictive of reoffending than having significant problems (AUC scores of .705 and .680 respectively). However, neither model was as predictive of reoffending as the static risk OGRS3 predictor (which had an AUC score of .776).

Table 6.7 also sets out the percentages correctly predicted for each model. As shown, the final regression model, combining promotive and protective factors with static risk and dynamic risk factors, produced a correct prediction in approximately seven out of ten (71%) of the cases in the validation sample.[93] The combined static risk and dynamic risk model performed almost as well. The lowest percentage correctly predicted was achieved by the model based purely upon dynamic risk factors, with a correct prediction in less than two-thirds (64%) of the cases.

Table 6.7: Comparative accuracy of logistic regression models

Model	AUC score	SE	95% confidence intervals		Percent correctly predicted	Residual (actual minus predicted rate)
			Lower bound	Upper bound		
Dynamic risk factors model	.680	.003	.675	.686	64.0%	0.2%
Positive factors model	.705	.003	.700	.710	65.8%	0.3%
Static risk factors (OGRS3)	.776	.002	.771	.781	70.7%	0.3%
Combined static risk and dynamic risk factors model	.781	.002	.776	.786	71.0%	0.3%
Combined static risk, dynamic risk and promotive factors model	.783	.002	.778	.788	71.2%	0.3%
Combined static risk, dynamic risk, promotive and protective factors model	.783	.002	.778	.787	71.2%	0.3%

92 Further checks on the goodness-of-fit of this model are set out in Appendix E.

93 Copas (1992, unpublished) explains that for an actual reconviction rate of 50 per cent, the proportion correctly predicted cannot normally exceed 75 per cent, even for an optimally effective predictor.

6.4 Implications

The textual analysis revealed that the positive factors recorded by assessors within the OASys sentence plan correspond to the socio-economic and individual-level domains covered by the core OASys assessment. Trichotomising the ten socio-economic and individual-level scales revealed that all ten domains were significantly associated with reoffending at the promotive ends of their scales and all except emotional wellbeing were significantly associated with reoffending at the risk ends of their scales. The optimum logistic regression model for predicting reoffending included the static risk OGRS3 score, six dynamic risk factors, five promotive factors and one protective interaction. However, this model performed only marginally better than a model combining static and dynamic risk factors alone.

The general implications for policy makers and practitioners are as follows:

- Promotive factors can exist despite the presence of numerous risk factors, and the identification of overlapping risk and promotive processes indicates that attempts should be made to strengthen promotive factors as well as reduce known risk factors. Where risk factors are hard to change, interventions can potentially offset the risks of further offending by enhancing promotive factors, assisting with offender engagement.
- Analysis of the interactions between factors indicates that positive family relationships can moderate the impact of problematic drug misuse. When addressing problematic drug misuse, attention should thus also be paid to maximising positive family relationships.

More specific implications for the development of OASys are as follows:

- The importance of identifying positive as well as risk factors should be highlighted further during assessors' training and their recording should be carefully monitored through existing quality assurance procedures.[94]
- The prevalence rates of the extracted positive factor categories (from the text field within the sentence plan) were relatively low, indicating that the full range of positive factors may not always have been considered. More systematic recording within the sentence plan could be achieved through the introduction of fixed response categories which should be ticked where appropriate. Distinguishing between positive factors that need to be maintained and those that need to be developed would assist in the identification of immediately promotive/protective factors and enable changes in status (development vs. maintenance) to be monitored.[95]

[94] The OASys Quality Assurance Tool, implemented in January 2010, requires designated quality assurers to consider whether positive and negative factors have been identified across each of the sections in the core OASys assessment.

[95] Such an approach is adopted within the SAPROF (Structured Assessment of Protective Factors for Violence Risk) checklist, which enables practitioners to distinguish between protective effects that are already present and those that are dependent upon intervention (de Vogel *et al*., 2008).

- The OASys predictors of general reoffending (OGP) and violent reoffending (OVP) have high predictive validity and this chapter suggests that little would be gained, in terms of accurately predicting reoffending, from a scoring system distinguishing risk factors from promotive/protective factors.

7. Reliability and validity of the risk of serious harm ratings

This chapter presents an analysis of the reliability and validity of OASys Risk of Serious Harm (RoSH) ratings. Probation assessments completed between 2005 and 2008 were analysed and those completed by mid-July 2006 were matched with 24-month reoffending data. Key findings were as follows:

- The analysis revealed not only variation between probation areas in high/very high RoSH prevalence rates, but also differences in the actual minus predicted residual rates. The majority of probation areas had significantly fewer high/very high risk offenders than predicted, with some large urban areas having more high/very risk offenders than predicted.
- There was also considerable variation between probation areas in the use of the RoSH screening overrides. Greater consistency could be encouraged through improved guidance regarding the use of the overrides and possibly the introduction of structured response options.
- Looking at half-year periods from 2005 to 2008, the increase in high/very high RoSH ratings appeared broadly justified by the characteristics of the offenders who were assessed. The more sizeable shift was from low to medium RoSH ratings.
- Grave reoffences were predicted with much greatly validity by an actuarial risk assessment score than by the clinical RoSH ratings. It is therefore likely that public protection could be improved by increasing the influence of actuarial scores upon RoSH ratings. As highlighted in Chapter 13, this has led to the development of a new actuarial Risk of Serious Recidivism (RSR) tool which is being used alongside the RoSH ratings in the Case Allocation System (CAS) for routing cases to the National Probation Service (NPS) or to Community Rehabilitation Companies (CRCs).

7.1 Context

The Risk of Serious Harm (RoSH) component of OASys is divided into a screening, full analysis and a summary. The screening is completed in all cases and is used to indicate whether the full analysis should be completed, with the summary drawing together the information from the previous sections. Nine ratings are determined through structured professional judgement: the risk to children, known adults, the general public and staff are rated for both the community and custodial settings, while the risk to other prisoners is rated for the custodial setting. Each of these risks are rated as low, medium, high or very high, with the standard summary measure in risk/need profile reports being the ‘highest community risk’ (representing the highest of the four ratings in the community setting). The accompanying guidance defines serious harm and the differing levels as follows:

> **Serious harm** can be defined as an event which is life-threatening and/or traumatic, and from which recovery, whether physical or psychological, can be expected to be difficult or impossible. The levels of RoSH used in OASys are:

Low – current evidence does not indicate likelihood of causing serious harm.
Medium – there are identifiable indicators of RoSH. The offender has the potential to cause serious harm but is unlikely to do so unless there is a change in circumstances, for example, failure to take medication, loss of accommodation, relationship breakdown, drug or alcohol misuse.
High – there are identifiable indicators of RoSH. The potential event could happen at any time and the impact would be serious.
Very high – there is an imminent RoSH. The potential event is more likely than not to happen imminently and the impact would be serious.

Importantly, the value of the RoSH ratings are dependent upon their ***reliability*** and ***validity***. For the ratings to be reliable, they need to be consistent, and to be valid they need to be measuring what they are intended to measure, i.e. the likelihood of serious harmful reoffending. To date, research on RoSH reliability and validity has been limited, although Morton's OASys inter-rater reliability study (Morton, 2009a) reported poor consistency. Two of the three case studies used in the study (overall n= 178) had poor consensus as to whether it was necessary to complete the full RoSH analysis and to the highest level of risk posed by the offender.

The central collation of OASys and reoffending data provide an opportunity to address the shortfall in findings on RoSH reliability and validity. Much of the focus in this chapter is upon those offenders rated as presenting high/very high RoSH to the community. Offenders rated as high/very high risk are often subject to Multi-Agency Public Protection Arrangements (MAPPA)[96] and are the target of more intensive supervision within the Offender Management model. Looking forward, as community rehabilitation services are opened up to a diverse range of new providers, the National Probation Service will remain responsible for the direct management of those offenders who pose the highest RoSH to the public (Ministry of Justice, 2013b; National Offender Management Service, 2014). Dealing with such cases demands considerable resources, and the decision to rate an offender as high/very high risk should therefore be made carefully and be the subject of appropriate scrutiny.

To guide the analysis reported in this chapter, the following four research questions were set:

1. Do the RoSH ratings differ between probation areas,[97] comparing practitioners' actual ratings to predicted ratings?
2. Does the use of the RoSH screening overrides (requiring or exempting the full analysis) differ between probation areas?

96 Although the levels of risk do not equate directly to the three levels of MAPPA management. The MAPPA guidance states as follows: "The central question in determining the correct MAPPA level is: 'What is the lowest level that a case can be managed at which provides a defensible Risk Management Plan?' This means that not all high-risk cases will need to be managed at level 2 or 3. Similarly, the complexities of managing a low/medium risk case might, in exceptional circumstances, justify it being managed at level 2 or 3, especially where notoriety is an issue." (National Offender Management Service, 2009:91).

97 For the period covered by the research, data was available for 42 probation areas. These areas have since been replaced by 35 probation trusts.

3. Have RoSH ratings changed over time, comparing practitioners' actual ratings to predicted ratings?
4. How does the predictive validity of the RoSH ratings compare to an actuarial predictor for grave offences?

The fourth research question assesses the extent to which the clinical ratings of offenders' RoSH levels provide foresight as to those most likely to commit serious offences in the future, assuming that any preventative actions are unsuccessful. The alternative is to rely upon actuarially scored predictors of reoffending, which combine a range of risk factors through a predetermined procedure to determine the offenders' risk scores or categories. Actuarial predictors can incorporate a narrow (e.g. OGRS: see Chapter 8) or broad (e.g. OGP and OVP: see Chapter 9) range of risk factors. The clinical ratings of RoSH within OASys also consider who is at risk and when this is likely to be greatest, but in order to add further value to the overall OASys assessment, they should ideally demonstrate greater predictive validity than the available actuarial predictors.

7.2 Approach

Sample

Data was taken from the O-DEAT database of completed OASys assessments. For the first three research questions, assessments were selected if they were completed by the probation service between 2005 and 2008 and satisfied O-DEAT's data completeness and de-duplication filters.[98] Some 516,461 assessments were selected, with more of these being completed in recent years due to improvements in the levels of OASys coverage and data completeness – an increase from 44,012 in the first half of 2005 to 125,690 in the second half of 2008. The RoSH full analysis was completed in 72% of the cases, with a final RoSH breakdown as follows: 43% low, 48% medium, 8.3% high and 0.3% very high. When exploring differences between probation areas (questions 1 and 2), those assessments completed during the final half-year of the sampling period (July to December 2008 were used, providing a sufficiently large sample (n=125,690) over the most recent time period available.

[98] The assessments were cleansed and de-duplicated by selecting valid assessments and prioritising the earliest such assessments in each individual contact period. For an OASys assessment to be held valid, the following standards of data completion had to be satisfied: (i) Each of the scored sections (1 to 12) within the core OASys assessment must have had at least four-fifths of their scored items completed – ensuring that each criminogenic need was assessed properly; and (ii) in the RoSH sections, the screening must have been completed, the decision whether to complete a full risk analysis should have been consistent with the information provided, and the four ratings of risk of serious harm in the community must have been recorded in those cases in which a full analysis was required.

For the fourth research question (predictive validity), OASys assessments completed between January 2002 and March 2008 were again filtered for data quality and de-duplicated. The selected offenders were traced on the Ministry of Justice's (MoJ's) Police National Computer (PNC) research database to ascertain their criminal history and proven reoffending rates, using a 24-month follow-up period and a subsequent six-month data entry 'buffer' period. Assessments were only retained where the two sources agreed upon the offender's age, gender and index offence conviction date. Having also excluded those offenders imprisoned for an offence committed before the start of follow-up (a 'pseudoreconviction'), and those recalled to custody before any grave reoffending (as the recall causes them to have an incomplete reoffending follow-up), 205,448 assessments remained for use in the analysis. The RoSH full analysis was completed in 67% of these cases, with a final RoSH breakdown as follows: 47% low, 46% medium, 6.5% high and 0.2% very high.[99]

To examine 'pure' prediction (see below), those assessments for offenders serving sentences without supervision or interventions were selected. These sentences were Community Punishment Orders (prior to the April 2005 implementation of the Criminal Justice Act 2003) and Community Orders or Suspended Sentence Orders with unpaid work, curfew, prohibited activity, exclusion and/or attendance centre requirements, and without any other requirements. There were 42,631 cases after excluding for pseudoreconviction and recall. The RoSH full analysis was completed in 50% of these cases, with a final RoSH breakdown as follows: 65% low, 34% medium, 1.3% high and less than 0.1% very high. Not surprisingly, the offenders were less likely than those in the full OASys/PNC sample to have been assessed as high/very high RoSH, although there remained 560 such cases for use in the analysis.

Analysis

Probation area level differences in RoSH ratings

To examine the extent to which differences in RoSH ratings between probation areas were linked to caseload differences, the analysis utilised a ten question checklist which had been developed previously (August 2009) to identify offenders likely to be rated as high/very high RoSH (assisting practitioners in deciding when a 15-day adjourned Pre-Sentence Report was most necessary – see Appendix F). Table 7.1 sets out the distribution of the offenders' high/very high RoSH checklist scores (for all the 2005 – 2008 assessments), and how they related to the practitioners' actual high/very high RoSH ratings.

99 The matched OASys/PNC sample thus had fewer high/very high RoSH offenders than the 2005-08 sample. This is partly due to an increase in RoSH rates over time (as set out in the results of this paper) - the matched sample had no cases after March 2008 and did have some cases prior to 2005 – and partly due to the removal from the matched sample of those recalled to custody during the 24-month follow up before any grave reoffending had occurred.

Table 7.1: High/very high risk of serious harm prevalence rate by 10-item checklist score

Score	Offenders with this score		% High/very high RoSH
	No.	%	
0	140,369	27.2	0.3
1	135,250	26.2	1.4
2	110,031	21.3	6.4
3	74,064	14.3	16.8
4	36,702	7.1	31.8
5	14,262	2.8	49.1
6+	5,783	1.1	69.8
All offenders	***516,461***	***100%***	***8.6***

The predictive validity of the checklist scores was found to be high, with an Area Under Curve (AUC) statistic of 0.867 – this is a higher AUC than found in O-DEAT's models for predicting reoffending (see Chapters 8 and 9), indicating that high/very high RoSH classification decisions are relatively predictable. However, the checklist appeared to work less well for female offenders (2.8% of female offenders were high/very high risk, compared with a 4.6% mean predicted rate) and older offenders (those aged 65 to 81 were at least 20% high/very high risk – over 30% at some ages – compared with mean predicted rates several points lower). The model used in this chapter for predicting high/very high RoSH was thus adapted to include a gender term and simple and quadratic age terms. The AUC for this modified model was 0.870 (see Appendix G for logistic regression model).

For the July to December 2008 assessments, the mean actual (i.e. practitioner assessed) and predicted rates of high/very high RoSH were then compared across the 42 probation areas. Predicted and actual rates were plotted for all areas, identifying areas with high and low levels of risk-related offender characteristics (those used in the checklist), actual risk rates, and above- and below-expected proportions rated high/very high RoSH. Further attention was given to those areas with the largest residuals (i.e. actual rates well above or below their predicted rates), examining the relationship between each of the ten checklist risk factors and high/very high risk status.

Probation area level differences in the use of RoSH screening overrides

For the July to December 2008 assessments, area variation in the use of the following two RoSH screening overrides was examined:

- R5.1: Is there anything else about the offender that leads you to consider that a full analysis should be completed.
- R5.2: *If you have ticked YES to any above you must complete the full analysis unless, in your judgement, there is a sound reason for not doing so.*

Areas were also compared by the residuals calculated above, assessing: (i) whether areas with the highest residuals (i.e. more high/very high risk offenders than predicted) were making more use of

R5.1; and (ii) whether areas with the lowest residuals (i.e. fewer high/very high risk offenders than predicted) were making more use of R5.2.

Changes in RoSH levels over time

The mean actual and predicted rates of high/very high RoSH were compared at a national level for each half-year time period between 2005 and 2008.[100] Changes in the mean predicted rates of high/very high RoSH could be due to changes in the distribution of any of the ten checklist risk factors, age or gender between 2005 and 2008. These changes were therefore also tracked over the eight half-years.

Predictive validity of RoSH ratings compared to an actuarial predictor

For all offenders in the matched OASys/PNC sample, scores were calculated on the OASys Violence Predictor v.1 (OVP1: Howard, 2009). OVP1 predicts reoffending over a broad range of violence-related offences, but has also been validated as a predictor of the most serious violent offences (i.e. homicide and wounding with intent to cause grievous bodily harm).

The predictive validity of the RoSH ratings were compared with OVP1 for 'grave' reoffences, covering homicide, attempted murder, wounding, rape, arson, robbery and aggravated burglary (Coid *et al.*, 2009). These offence types are covered by R1.2 (*Has the offender ever been convicted of any of the following (serious)* offences) of the RoSH screening, requiring the full analysis to be usually completed. The comparison was conducted for both the whole sample and those offenders serving sentences without supervision or interventions. The latter subsample was used to examine 'pure' prediction, checking whether (for the whole sample) an effective concentration of supervision and intervention resources upon offenders rated as high/very high RoSH had consequently reduced their actual levels of reoffending, thus leading to an underestimation of the predictive validity of the RoSH ratings. The same process would not apply to OVP1 scores, as OVP1 was not implemented until August 2009. For the explanation to hold true, any differences between the predictive validities of OVP1 and the RoSH ratings would be greater for the non-supervised offenders than the whole sample, due to the RoSH ratings not being acted upon in the former cases.

The comparisons of predictive validity were conducted using AUC statistics, the standard measure of such validity. AUC scores are generally higher for continuous predictors such as OVP1, as the calculation method favourably separates scaled scores which could have been tied in a grouped predictor such as the RoSH rating. Therefore, a fair comparison was ensured by banding ranked OVP1 scores into groups containing numbers of offenders corresponding to those in each RoSH category (low, medium, high and very high). These bands were created separately for the whole sample and the non-supervised subsample.

---

[100] Regional and probation area changes over time were also calculated, but the detailed tables are not presented in this chapter.

7.3 Results

Probation area level differences in risk of serious harm ratings

Focusing upon those 125,690 assessments completed during the second half of 2008, there were considerable differences between the 42 probation areas in actual high/very high RoSH rates and even greater differences in predicted rates. In Table7.2, the areas are ordered by their residuals (actual minus predicted rates), with the most negative first.[101] At the extremes, Lincolnshire assessed 4.4% fewer of their offenders as high/very high risk than predicted, while Merseyside assessed 4.2% more of their offenders as high/very high risk than predicted. The residuals do not always follow the order of the actual rates; for example, Suffolk had the third lowest residual, but its actual rate of 7.7% was only twelfth lowest. The majority of areas had significantly negative residuals. This apparent imbalance occurred because most of these areas had smaller caseloads. The large urban areas of Merseyside, London, Greater Manchester and West and South Yorkshire all had substantial positive residuals, with actual rates of at least 12% high/very high risk.

When areas' percentages on each of the ten checklist items were compared with their residuals, the (i) murder/manslaughter/GBH/wounding/robbery and (ii) weapon carrying/use percentages were strongly positively correlated with the residuals (and excessive/sadistic violence was slightly less strongly correlated). That is, those areas with many offenders who had a record of serious violent offences and/or who carried/used weapons in the current offence were more likely to have a greater proportion of high/very high risk offenders than predicted (even though, at an individual offender level, these three items were not exceptionally strong predictors of high/very high RoSH status). This supports the view that areas with a 'tough' caseload, as reflected in the propensity of its offenders to be involved in serious non-domestic, non-sexual violence, were more likely to classify offenders as higher risk.

[101] Table 7.2 demonstrates that there was also substantial diversity in allocation to the low and medium RoSH categories. The table includes an 'average' highest community risk statistic. This is derived by scoring low risk as 1, medium risk as 2, high risk as 3 and very high risk as 4. It is only useful for indicative purposes as (for example) a very high risk offender cannot meaningfully be described as four times as harmful as a low risk offender.

Table 7.2: Actual and predicted risk of serious harm rates by probation area (ranked by residuals)

Probation area	n	Highest RoSH in the community					Actual and predicted high/very high RoSH rates		
		% low	% medium	% high	% very high	Average *(L=1, M=2 H=3,VH=4)*	Actual %	Predicted %	Residual % *(Actual Minus predicted)*
Probation areas in which actual high/very high RoSH rate was significantly lower than predicted (p=.05)									
Lincolnshire	1,194	54.0	40.6	5.2	0.2	1.52	5.4	9.8	-4.4
Sussex	2,676	42.9	50.6	6.2	0.3	1.64	6.5	10.5	-4.0
Suffolk	1,237	38.3	54.0	7.4	0.2	1.70	7.7	11.5	-3.8
Teesside	2,782	45.3	49.8	4.6	0.3	1.60	4.9	8.6	-3.7
Gwent	1,880	41.4	52.7	5.8	0.1	1.65	5.9	9.5	-3.6
Durham	2,018	33.5	60.9	5.6	0.0	1.72	5.6	9.0	-3.5
Devon and Cornwall	2,488	38.2	53.6	7.8	0.4	1.70	8.2	11.0	-2.9
Hertfordshire	2,104	54.6	40.8	4.5	0.1	1.50	4.7	7.3	-2.7
Leicestershire	2,613	37.0	57.7	5.1	0.2	1.69	5.3	7.8	-2.5
Wiltshire	1,149	29.1	62.5	8.3	0.2	1.80	8.4	10.7	-2.3
Cambridgeshire	1,454	33.3	59.1	7.2	0.4	1.75	7.6	9.8	-2.2
Hampshire	4,004	33.5	57.9	8.5	0.2	1.75	8.7	10.7	-2.1
Thames Valley	3,973	39.3	52.9	7.6	0.2	1.69	7.8	9.9	-2.1
Dorset	1,522	48.6	44.0	7.3	0.1	1.59	7.4	9.2	-1.7
Norfolk	1,217	33.7	55.2	10.8	0.2	1.78	11.1	12.6	-1.5
Bedfordshire	1,536	54.3	39.3	6.2	0.2	1.52	6.4	7.6	-1.2
Dyfed and Powys	1,123	47.3	45.4	6.9	0.4	1.60	7.3	8.3	-1.0
Cheshire	2,274	22.5	68.0	9.3	0.2	1.87	9.5	10.5	-1.0
Gloucestershire	1,386	30.0	61.2	8.3	0.5	1.79	8.8	9.8	-1.0
Northumbria	4,127	38.8	52.8	8.0	0.5	1.70	8.4	9.4	-0.9
Avon and Somerset	3,065	24.8	64.6	10.3	0.2	1.86	10.5	11.4	-0.9
Cumbria	1,159	43.1	48.3	8.5	0.2	1.66	8.6	9.5	-0.9

Probation area	n	Highest RoSH in the community					Actual and predicted high/very high RoSH rates		
		% low	% medium	% high	% very high	Average *(L=1, M=2 H=3,VH=4)*	Actual %	Predicted %	Residual % *(Actual Minus predicted)*
Nottinghamshire	3,005	35.7	55.2	8.4	0.7	1.74	9.1	9.8	-0.7
Humberside	2,533	40.3	48.6	10.9	0.2	1.71	11.1	11.7	-0.6
Derbyshire	2,435	27.6	62.6	9.5	0.2	1.82	9.8	10.2	-0.4
West Midlands	9,738	41.4	50.1	8.2	0.3	1.67	8.5	8.8	-0.3
Probation areas in which actual high/very high RoSH rate was not significantly different to predicted rate (p=.05)									
Kent	3,075	42.9	47.4	9.3	0.4	1.67	9.7	10.0	-0.3
North Wales	1,382	23.1	66.1	10.3	0.5	1.88	10.9	11.1	-0.2
Northamptonshire	1,683	30.5	59.7	9.5	0.2	1.79	9.7	9.9	-0.2
Lancashire	4,213	42.9	48.1	8.8	0.1	1.66	9.0	9.1	-0.1
West Mercia	1,938	36.9	53.2	9.5	0.4	1.73	9.9	9.7	0.2
North Yorkshire	1,454	52.1	39.1	8.5	0.3	1.57	8.8	8.5	0.3
Probation areas in which actual high/very high RoSH rate was significantly higher than predicted (p=.05)									
Essex	3,002	34.4	56.4	9.0	0.2	1.75	9.2	8.8	0.4
South Wales	3,263	34.1	54.8	10.6	0.5	1.77	11.1	10.6	0.4
Surrey	1,231	41.3	48.0	10.6	0.2	1.70	10.7	9.5	1.2
West Yorkshire	5,560	28.6	58.7	12.4	0.4	1.84	12.7	11.1	1.6
South Yorkshire	3,078	43.6	44.4	11.8	0.3	1.69	12.1	10.2	1.8
Greater Manchester	9,313	30.6	56.4	12.6	0.5	1.83	13.0	10.8	2.2
Warwickshire	971	32.1	55.7	11.8	0.3	1.80	12.2	9.9	2.3
Staffordshire	2,476	32.8	54.6	12.3	0.4	1.80	12.6	9.6	3.0
London	14,262	27.6	58.9	13.1	0.3	1.86	13.5	10.1	3.3
Merseyside	4,518	38.3	49.2	12.1	0.4	1.75	12.5	8.2	4.2

Probation area level differences in the use of risk of serious harm screening overrides

Figure 7.1 shows area variation in the use of R5.1 (*Is there anything else about the offender that leads you to consider that a full analysis should be completed*), with overall rates ranging widely from 16% in Bedfordshire and Surrey to 56% in North Wales. In the high usage areas, practitioners were seemingly finding the preceding fixed response questions insufficient for reflecting the full range of potential RoSH issues. However, further analysis revealed that the full analysis was completed as a result of R5.1 alone, with no positive responses to the preceding questions (R1.1 to R4.4), in just two percent of the cases. When the use of R5.1 was categorised into low (<25%), medium (25%–34%), high (35%–39%) and very high (40%+), areas fell within the following categories (moving from left to right on Figure 7.1):

- Low: 8 areas (Bedfordshire to Leicestershire and Rutland)
- Medium: 19 areas (Cambridgeshire to Devon and Cornwall)
- High: 9 areas (Cheshire to Merseyside)
- Very high: 6 areas (Nottinghamshire to North Wales)

To identify the types of issues being recorded within R5.1, the textual data was analysed using a linguistic-based text mining tool.[102] Key concepts/terms, representing the essential information, were extracted automatically, with closely related concepts then grouped into higher-level categories, firstly through further linguistic based methods and then manually. The most prevalent such categories, all recorded in at least five percent of those assessments in which R5.1 was ticked, are set out below. Combining the categories, reference was made to an offence (previous or current) or to some form of violence (domestic or otherwise) in approximately half (51%) of the cases.

- Offence 28.7%
- Violence 26.2%
- Financial issues 16.5%
- Family 14.5%
- Convictions 9.2%
- Domestic violence 6.8%
- Weapons 5.7%
- Accommodation 5.5%
- Alcohol 5.1%

Similarly to R5.1, there was wide variation in the use of R5.2 (If you have ticked YES to any above you must complete the full analysis unless, in your judgement, there is a sound reason for not doing so),

[102] The tool (a module within IBM SPSS Modeller) employs advanced linguistic technologies and natural language processing. Normalisation and grouping techniques correct punctuation and spelling errors respectively, with linguistic methods identifying synonym and hyponym relationships and root terms.

with overall rates ranging from 2% in Hampshire to 26% in Humberside (see Figure 7.2).[103] The use of an R5.2 exemption has to be clearly evidenced, the assessor must be confident that the offender is not likely to cause serious harm and it has to be countersigned by a senior practitioner. An exemption rate of 26% may therefore be viewed as higher than expected. When the use of R5.2 was categorised into low (<5%), medium (5%–9%), high (10%–14%) and very high (15%+), areas fell within the following categories (moving from left to right on Figure 7.2):

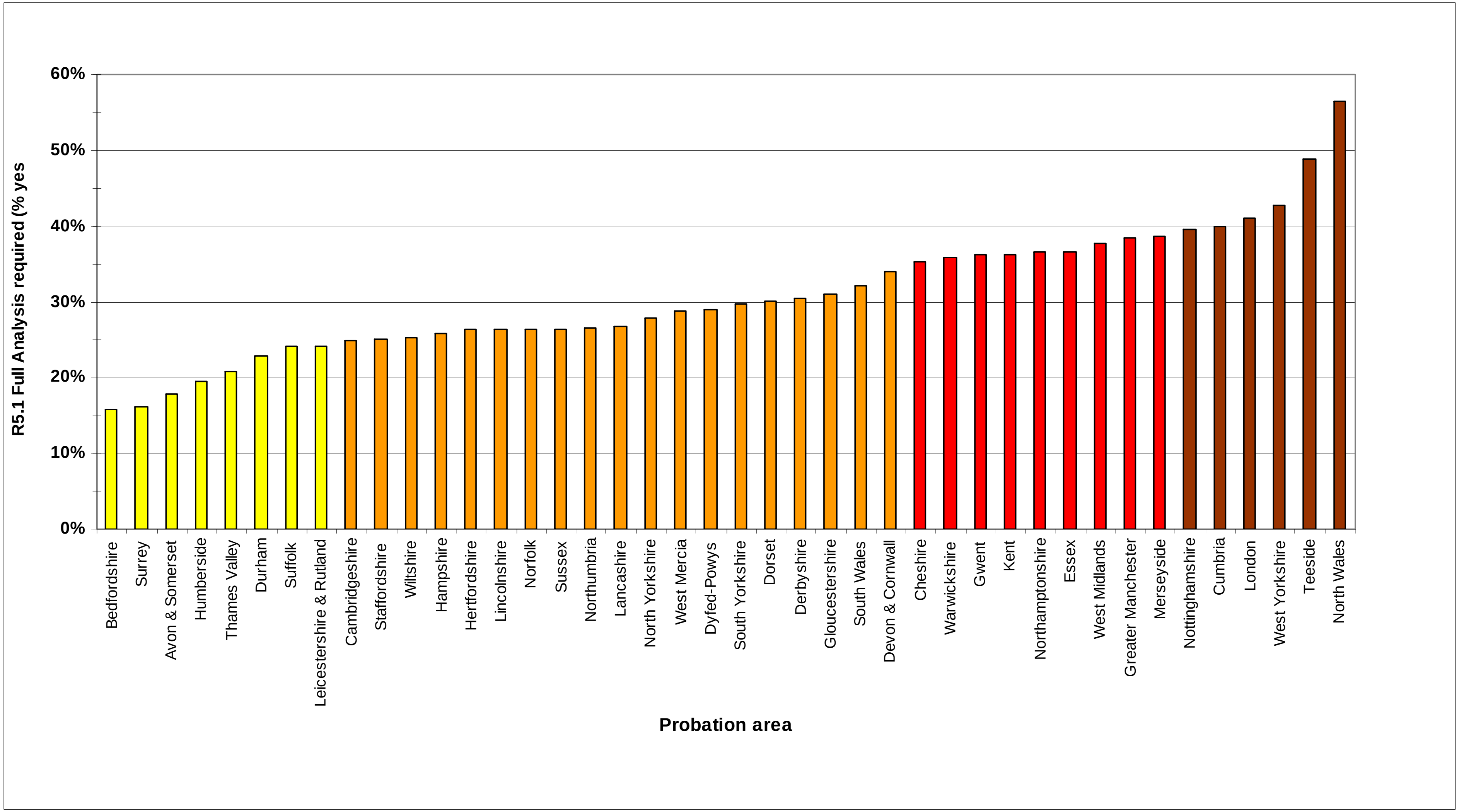
- Low: 7 areas (Hampshire to Northamptonshire)
- Medium: 16 areas (London to Lancashire)
- High: 12 areas (Suffolk to Hertfordshire)
- Very high: 7 areas (Kent to Humberside)

Further analysis revealed that five percent of the exemptions appeared to be unnecessary, with no positive responses to any of the preceding questions (R1.1 to R5.1). This 'false exemption' rate ranged from 2% in Humberside to 16% in Northamptonshire, but there was no clear relationship between these percentages and the overall prevalence rates in the use of R5.2. Text mining of the information recorded in those cases in which R5.2 had been ticked revealed that reference was made to the offence itself in half (50.4%) of the cases, with the offender's convictions being highlighted in over a quarter (28.7%) of the cases. Notable other extracted categories were education, training and employment (14.9%) and family members (13.6%), with the latter having a sub-category of children (10.0%). Specific reference was made to some type of 'change' in 6.8% of the cases.

Figure 7.3 shows areas ordered by their high/very high RoSH residuals (actual minus predicted rates), with the lowest (Lincolnshire) on the left to the highest (Merseyside) on the right of the chart. As can be seen, there was no clear pattern between the residual levels and the use of either R5.1 or R5.2. Looking at the lowest residual areas, Lincolnshire and Sussex differed greatly in their use of exemptions (R5.2) – 20% and 3% of the cases respectively. Similarly, looking at the highest residual areas, Staffordshire and Merseyside differed greatly in their use of R5.1 – 25% and 39% of the cases respectively.

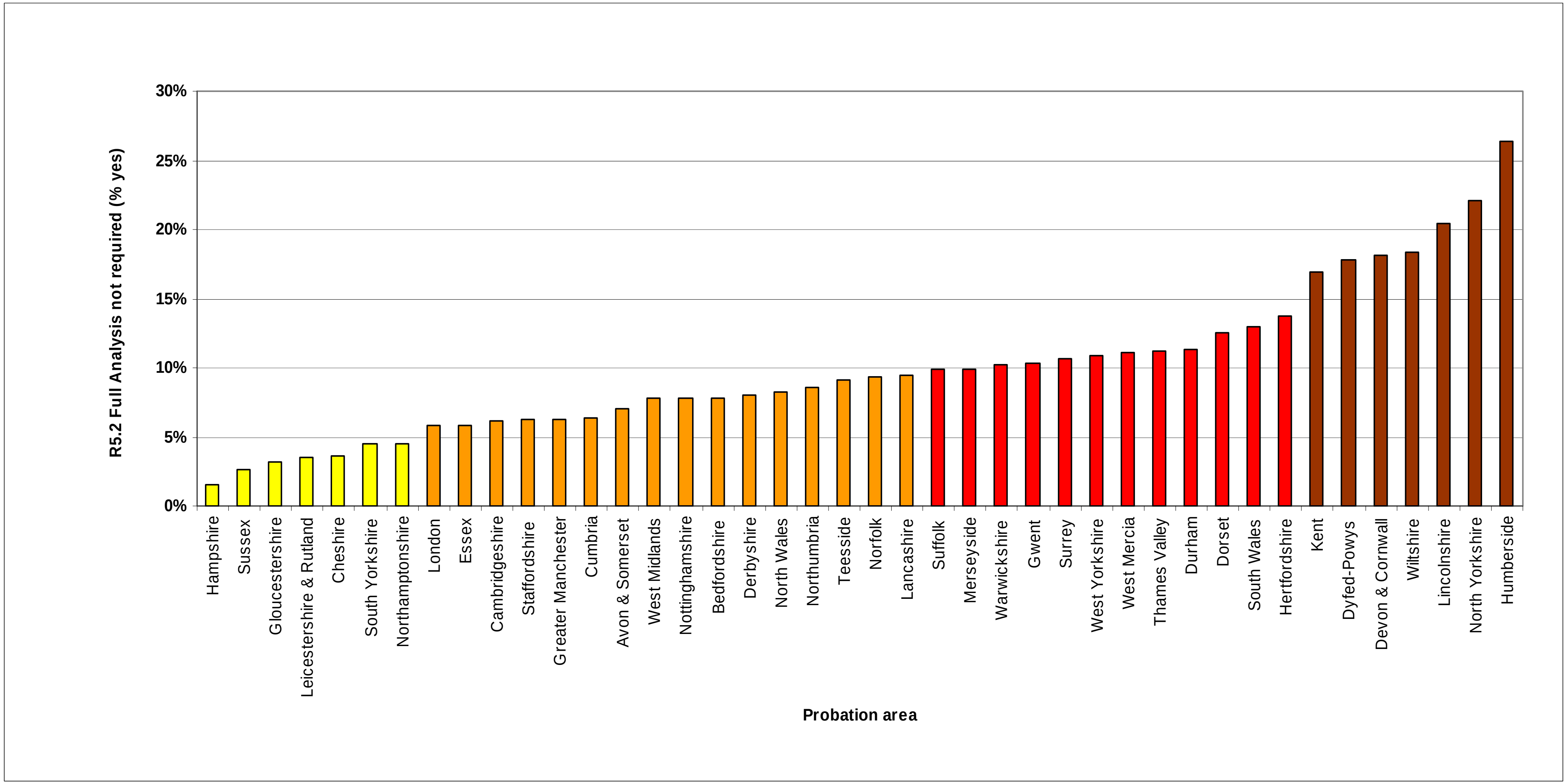
[103] In three percent of the assessments, R5.1 was ticked but an R5.2 exemption was used and the full analysis was not completed. There were no assessments in which an R5.2 exemption was used but the full analysis was still completed.

Figure 7.1: Use of R5.1 (full risk of serious harm analysis required) by probation area



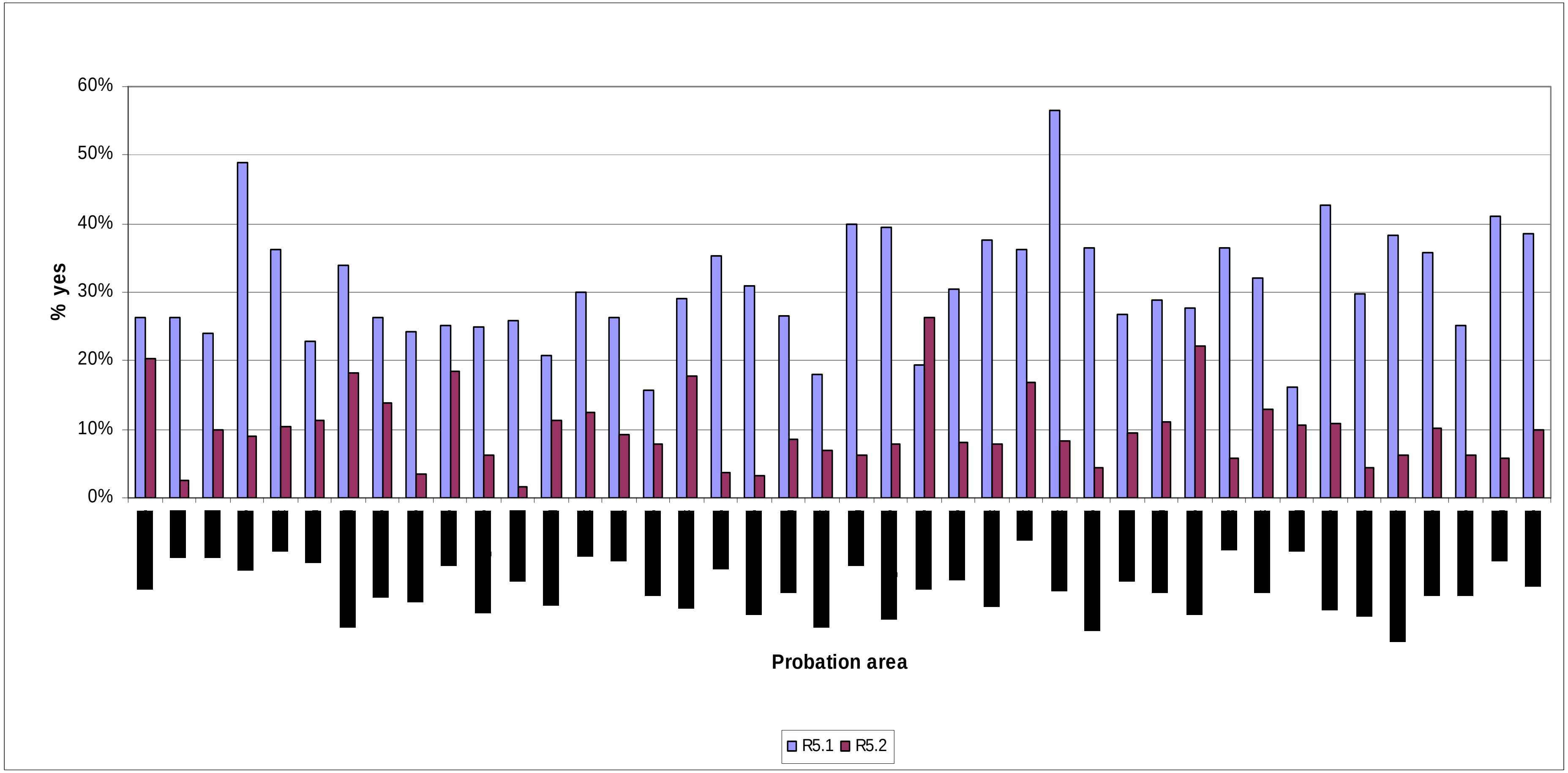
Key: % use of R5.1 Low Medium High Very high

Figure 7.2: Use of R5.2 (full risk of serious harm analysis not required i.e. exempted) by probation area



Key: % use of R5.2 Low Medium High Very high

Figure 7.3: Use of R5.1 and R5.2 by probation areas ranked by high/very high risk of serious harm residuals



Changes in risk of serious harm levels over time

Table 7.3 sets out RoSH distributions at the national level for the eight half-year periods between 2005 and 2008. The 'predicted' high/very high rate shows how the risk-relevant characteristics of the caseload's criminal history and current offences changed over time, and the 'residual' tracks the movement of the actual high/very high rates compared with the predicted high/very high rates. As can be seen, there was an increase in the proportion of offenders assessed as high/very high RoSH from 5.8% to 9.9%. Notably, the shift in ratings appears to have been broadly justified by the characteristics of the assessed caseload, as the residuals for each half-year period were small (although sometimes statistically significant).[104] The table also demonstrates a large shift from low to medium RoSH ratings; the former falling from 62% to 36% and the latter increasing from 32% to 54%. To some extent, this can be explained by the fact that since around 2006, the OASys eligibility guidance has emphasised that offenders at Offender Management Tier 1 should not usually be assessed. Under the tiering framework (National Offender Management Service, 2008), offenders within this lowest tier have to be low RoSH.

[104] In the large national sample, very small residuals can be statistically significant.

Table 7.3: Actual and predicted risk of serious harm rates over time

		Highest RoSH in the community					Actual and predicted high/very high RoSH rates			
Half year (H1 = Jan – June, H2 = July – Dec)	**n**	**% low**	**% medium**	**% high**	**% very high**	**Average** ***(L=1, M=2 H=3,VH=4)***	**Actual %**	**Predicted %**	**Residual %** ***(Actual minus predicted)***	**Significance of residual** ***(+ = sig. more than predicted, - = sig. less than predicted, [blank] = not sig. at p=.05)***
2005 H1	44,012	62.0	32.2	5.5	0.3	1.44	5.8	6.2	-0.5	**-**
2005 H2	41,217	53.5	39.2	7.0	0.4	1.54	7.3	7.4	0.1	
2006 H1	49,045	49.1	43.5	7.0	0.3	1.59	7.4	7.8	-0.4	**-**
2006 H2	51,965	45.4	47.2	7.1	0.3	1.62	7.3	7.8	-0.5	**-**
2007 H1	58,125	42.2	49.4	8.0	0.3	1.66	8.4	8.4	0.0	
2007 H2	63,824	40.2	50.5	8.9	0.4	1.69	9.3	9.0	0.2	**+**
2008 H1	82,583	37.2	52.7	9.7	0.4	1.73	10.1	9.7	0.3	**+**
2008 H2	125,690	36.1	54.1	9.6	0.3	1.74	9.9	9.7	0.2	**+**

Predictive validity of risk of serious harm ratings compared to an actuarial predictor

To examine the extent to which the clinical ratings of offenders' RoSH levels provided foresight as to those most likely to commit serious offences, Table 7.4 compares the RoSH ratings with banded OVP1 scores as predictors of proven 'grave' reoffending (for the matched OASys/PNC sample). As can be seen, OVP1 scores demonstrated substantially better predictive validity; the difference in AUCs (0.60 vs. 0.70) was highly significant (chi-square = 850.7, p<.0001).[105] The seven percent of offenders who were identified by OVP1 as high/very high risk offenders (basing this on the RoSH numbers, rather than boundaries set out in user guidance) comprised 24% of proven grave reoffenders. Using the clinical RoSH ratings, only 12% of such reoffenders had been subject to the enhanced supervision associated with being in the highest-risk groups. Conversely, 14% of such reoffenders were low risk under OVP1 compared with 30% under the RoSH ratings.

Table 7.4: Prediction of 24-month proven 'grave' reoffending (RoSH vs. OVP1): all offenders

RoSH rating	Equivalent OVP1 scores	Number of offenders (% of total N)	*n* of this risk group reoffending (% of this risk group reoffending; these reoffenders as a % of all such reoffenders)	
			RoSH (AUC = .60)	OVP1 (AUC = .70)
Low	0 – 38	96,271 (46.9%)	1,704 (1.8%; 29.8%)	809 (0.8%; 14.2%)
Medium	38 – 60	95,412 (46.4%)	3,317 (3.5%; 58.1%)	3,528 (3.7%; 61.8%)
High	60 – 77	13,341 (6.5%)	669 (5.0%; 11.7%)	1,287 (9.7%; 22.5%)
Very high	77 – 100	424 (0.2%)	23 (5.4%; 0.4%)	89 (21.0%; 1.6%)
Total		***205,448 (100%)***	***5,713 (2.8%; 100%)***	

Table 7.5 makes the same comparisons for those receiving community sentences without any supervision or intervention content.[106] As shown, there is little evidence that the predictive validity of the clinical RoSH ratings were improved by focusing on unsupervised offenders: neither AUC changed by more than .002, and the difference was again highly statistically significant (chi-square = 108.8, p<.0001).[107]

105 A further comparison was made using an alternative definition for "grave" reoffences covering homicide and wounding plus contact sexual offences. This definition included more sexual offences plus threats to kill, while omitting GBH-type wounding, arson, robbery and aggravated burglary. Using this definition, the AUC difference was 0.61 vs. 0.67, chi-sq=106.4, p<.0001.

106 While the distribution of RoSH ratings was far lower for this subsample, the threshold OVP1 scores were slightly higher (i.e. compared with the whole sample, it appears that the level of risk was slightly underestimated for individual offenders).

107 Using the alternative definition for "grave" reoffences covering homicide and wounding plus contact sexual offences, the AUC difference was .58 and .65, chi-square = 22.0, p<.0001.

Table 7.5: Prediction of 24-month proven grave reoffending (RoSH vs. OVP1): offenders without intervention or supervision

RoSH rating	Equivalent OVP1 scores	Number of offenders (% of total N)	*n* of this risk group reoffending (% of this risk group reoffending; these reoffenders as a % of all such reoffenders)	
			RoSH (AUC = .60)	OVP1 (AUC = .70)
Low	0 – 40	27,724 (65.0%)	371 (1.3%; 47.0%)	210 (0.8%; 26.6%)
Medium	40 – 64	14,347 (33.7%)	385 (2.7%; 48.7%)	528 (3.7%; 66.8%)
High	64 – 84	557 (1.3%)	34 (6.1%; 4.3%)	50 (9.0%; 6.3%)
Very high	85 – 100	3 (0.01%)	Zero	2 (66.7%; 0.3%)
Total		***42,631 (100%)***	***790 (1.9%; 100%)***	

7.4 Implications

The results of this research indicate that the increase in high/very high RoSH ratings between 2005 and 2008 were broadly justified by the characteristics of the offenders who were assessed – the more sizeable shift was from low to medium RoSH ratings. However, when comparing probation areas, there was variation not only in high/very high RoSH prevalence rates but also differences in the actual minus predicted residual rates. The majority of probation areas had significantly fewer high/very high risk offenders than predicted, with the large urban areas of Merseyside, London, Greater Manchester, West and South Yorkshire all having substantially more high/very high risk offenders than predicted. There was also considerable variation between probation areas in the use of the RoSH screening overrides. Notably, use of the exemption from full analysis clause ranged from 2% to 26%. Bearing in mind that (i) every exemption has to be clearly evidenced, (ii) the assessor must be confident that the offender is not likely to cause serious harm and (iii) the exemption has to be countersigned by a senior practitioner, the higher rate may be viewed as higher than expected. Greater consistency could be encouraged through improved guidance regarding the use of the overrides and possibly the introduction of structured response options.

Perhaps most importantly, the analysis found that grave reoffending was predicted with much greater validity by an actuarial risk assessment score than by the clinical RoSH ratings. It is therefore likely that public protection could be improved by increasing the influence of actuarial scores upon RoSH ratings, and, as highlighted in Chapter 13, this has led to the development of a new actuarial Risk of Serious Recidivism (RSR) tool (see Appendix H) which is being used alongside the RoSH ratings in the Case Allocation System (CAS) for routing cases to the National Probation Service (NPS) or to Community Rehabilitation Companies (CRCs). By structuring practitioners' judgements in this way, the intention is to ensure that the most appropriate high risk cases remain with the NPS.

8. OGRS4: the revised Offender Group Reconviction Scale

Chapters 8 to 10 of this compendium focus on the development and validation of new static and static/dynamic actuarial predictors of reoffending, covering general, violent and sexual reoffending. Findings from the earlier chapters feed into the development of these predictors. This chapter presents version 4 of the static predictor OGRS, setting out the following key points:

- OGRS4 includes models for general (i.e. all recordable) and violent proven reoffending, known as OGRS4/G and OGRS4/V respectively. In predicting general reoffending, OGRS4/G was found to significantly outperform OGRS3. In predicting violent reoffending, OGRS4/V significantly outperformed OGRS4/G and other operational predictors.
- The new models included an 'offence-free time' element, recognising that an offender's probability of future proven reoffending falls with time after community sentence or discharge from custody without yet reoffending (see Chapter 4). The models thus allow a more accurate comparison of offenders at different stages of community supervision, assisting with the targeting of supervision and treatment resources.
- The improvements in the prediction of both general and violent reoffending were due to the application of offence-free time and other innovations in the coding of risk factors. The choice of 'primary' static risk factors – those which must be entered by practitioners – was nevertheless constrained to ensure that all could be coded quickly. The refinements to the coding of the 'secondary' risk factors, calculated from the practitioner-entered information, illustrate the degree of fine-tuning required to achieve incremental improvements in the prediction of proven reoffending.
- The nature of the sample used to create OGRS4 means that the new predictors have scope to be used in settings where OGRS3 is not currently used, among offenders with cautions or absolute/conditional discharges from court, and in youth justice. Such use would require the development of user guidance and possibly training. Users who are already familiar with OGRS3 could be issued with more limited guidance covering the improved validity, revisions to offence categories, the offence-free time element, and any subsequent revisions to risk groupings.

8.1 Context

Those cautioned or convicted for criminal offences vary greatly in how likely they are to reoffend. Predictions of the likelihood of recidivism provide important information to those responsible for sentencing, assessing and treating these offenders, with the 'What Works' risk principle indicating that (rehabilitative) interventions should be targeted at moderate and high risk cases with low risk cases receiving minimal intervention The predictions also inform those commissioning and controlling offender management services, and aid research and evaluation of correctional services and interventions.

The Offender Group Reconviction Scale (OGRS) is the static, actuarial predictor used by the probation and prison services of England and Wales. Static actuarial predictors such as OGRS are based on a limited range of risk factors, such as age, gender and criminal history, although computerisation can allow these few factors to be scored in a sophisticated manner. It is therefore practical to use them in a wide range of situations, where it may not be possible to complete more thorough assessments encompassing socio-economic and personal risk factors. OGRS scores are used, or potentially can be used, to assess offenders who receive a non-court disposal such as a caution or minor court disposals such as absolute or conditional discharges from court, at pre-sentence court report stage, and post-sentence.

As demonstrated by Table 8.1, OGRS has been revised every few years. November 1996 saw the launch of the first version of OGRS (Copas and Marshall, 1998). It contained six simple demographic and criminal history factors, and was scored by probation officers using pencil, paper and calculator. As it was developed using the limited data available in the Home Office's Offenders Index (OI), both criminal history and the prediction of recidivism were limited to convictions for 'standard list' offences – this excluded some summary convictions and all non-conviction sanctions. In 2000, a revised version (OGRS2) was launched (Taylor, 1999). In an effort to improve prediction further, the factors included rose to ten. A computerised version was adopted, and OGRS was incorporated within OASys.
A separate predictor of violent and sexual reconviction was launched.

Version 3 was introduced into probation practice in February 2008 and prison practice in August 2009 (Howard *et al*., 2009). In developing OGRS3, a stronger emphasis was placed on reducing data coding and entry burdens. The number of factors was reduced to seven, while still improving the validity of predictions. As well as the existing two-year predictor, OGRS3 included a predictor of proven reoffending within one year of discharge or start of Community Order. The data source changed, becoming the Police National Computer (PNC), which includes all summary offences and non-conviction sanctions for recordable offences. This made OGRS3 more comprehensive, especially for young offenders, who were more likely to receive the non-conviction sanctions of reprimand and final warning. OGRS3 was therefore described as a predictor of 'proven reoffending' rather than 'reconviction for standard list offences'. In order to focus more on offending than the functioning of the criminal justice system, the one or two year follow-up period was based on the date of first proven reoffending rather than the date of reconviction. The violent and sexual reconviction predictor was not updated in OGRS3.

Table 8.1: OGRS versions 1 to 3

	Version		
	1	**2**	**3**
Sample and implementation			
Year offenders were sentenced/released	1990	1995	2002
Number of cases	14,000	30,000	79,000
Year implemented	1996	2000	2008/9

	Version		
	1	2	3
Factors included in the model:			
Age and gender			
Gender	√	√	(AG)
Age at time of sentence	√, (C)	√, (C)	(C)
Age at release or start of order			(AG)
Combination of age and gender			√
Age at first conviction	(C)	√, (C)	(C)
Offence/offending history			
Type of offence (number of categories)	√ (9)	√ (27)	√ (20)
Number of previous convictions	(C)	(C)	
Current or previous breach		√	
Current or previous burglary		√	
Number of previous youth custodial sentences	√	√	
Number of previous sanctions (convictions and cautions/reprimands/final warnings (CRFW))			(C)
Offending history status (first conviction; other conviction; first CRFW; second CRFW, or other CRFW)			√
Is current sanction a conviction or another sanction?			(O)
'Copas rate'[108]	√	√	√

Key: √ Included in its own right; (AG) Part of age/gender: (C) Part of 'Copas rate'; (O) Part of offending history status.

This chapter reports on the development of OGRS version 4 (OGRS4), which combines static and time-dependent risk factors (see next section). There are several reasons why actuarial risk assessment instruments should be revised periodically, and in particular why OGRS4 was created.

- Patterns of reoffending can change over time, resulting in increasing numbers of over- or under-predictions being made as the predictor 'ages' (as shown in Chapter 3). Instruments which produce precise predictions, as OGRS does, should at very least be recalibrated every few years.
- Various insights have been made since the development of OGRS3, which could improve prediction. Particularly notable was the realisation that an offender's 'hazard' (probability) of future proven reoffending falls with time after community sentence or discharge from custody without yet reoffending (Howard, 2011). The current use, for probation sentence planning, of a fixed OGRS3 score throughout the period after sentence implicitly assumes that this probability stays constant. On a practical level, taking account of offence-free time could enable changes in resourcing levels over the course of the sentence, and

[108] The distinction between age at sentence and age at release or start of order was clarified during the development of OGRS3. It was recognised that where an offender had served a sentence in custody for a non-trivial length of time, the choice of age could make a considerable difference to the prediction. As the Copas rate looks at past behaviour whereas age is based on the offender's current status, it was deemed appropriate to base the Copas rate on age at sentence but to base the age/gender risk factor on age at release.

allow more accurate comparison of offenders at different stages of community supervision when allocating scarce intervention places.

- OGRS3 lacks a predictor of violent recidivism, yet it has often been used to predict this outcome in forensic mental health research and practice (Yang, Wong and Coid, 2010). The recent direction of NOMS' assessment policy[109] has created a potential need for a cost-effective predictor of non-sexual violent recidivism. The only available such predictor, Risk Matrix 2000/v (RM2000/v), was designed for use with sex offenders, and has to be scored manually. While it has been used well with other offender groups, its simple scoring algorithm predicts violent recidivism no better than the OGRS3 score (Yang, Liu and Coid, 2010; Howard and Dixon, 2011).
- While OGRS3 was created for use with all offenders above the age of criminal responsibility, it was not implemented for juvenile offenders. Recent youth justice research (Wilson and Hinks, 2011) suggests that its predictive validity is moderate in absolute terms but good compared with other available scores for this group. The efforts set out in this chapter to predict well for offenders with limited criminal history may therefore be timely.

The aims of the analysis in this chapter were therefore:

1. To explore a recent dataset of offenders to confirm that accounting for offence-free time improves prediction.
2. To examine which static and offence-free time risk factors were most helpful in predicting all and non-sexual[110] violent proven reoffending.
3. To develop OGRS4 models for general (i.e. all recordable) and violent proven reoffending (OGRS4/G and OGRS4/V). These models were designed to predict 'next two years' reoffending: reoffending either in the two years following community disposal or discharge from custody, or in the two years following a subsequent offence-free period ranging from one month to three years.
4. To validate the new predictors on a further sample of offenders, comparing them with OGRS3 and the static part of the OASys Violence Predictor v.1 (OVP1; Howard, 2009), and checking their validity for offenders of different age, gender and ethnicity.[111]

The focus on 'next two years' reoffending aims to correctly prioritise offenders with different amounts of offence-free time, unlike the wholly static, 'first two years', OGRS3 scores. With the exception of offence-free time, the risk factors used should be similar to those used in OGRS3, to avoid increasing the resource cost of static risk assessment. Like previous versions, the risk scores should be applicable to offenders given the complete range of disposals, ranging from cautions, fines and

---

109 To ensure that all offenders are consistently assessed, with more in-depth assessments reserved for higher risk offenders.
110 NOMS currently uses the static risk predictor Risk Matrix 2000/S (Thornton, 2007) to assess proven sexual reoffending risk among offenders with a known history of sexual offending (see Chapter 10).
111 The validity of OGRS4/V was also checked for the outcome of proven homicide/wounding reoffending, ensuring that it improves prediction of the most harmful non-sexual violent recidivism.

discharges to the community-based orders and custodial sentences. It should cover juvenile as well as adult offenders.[112]

8.2 Approach

Sample

Traditional reoffending follow-ups start on the day of an offender's conviction leading to a community sentence or upon discharge from custody. A two-year follow-up period is standard, partly through historical precedent (Copas and Marshall, 1998) and partly as most community sentences last either one or two years (Ministry of Justice, 2010b). In this report, we refer to such traditional follow-ups and the predictions produced using them as '***first-two-year***' follow-ups or predicted rates, as they are based on the first two years in which offenders are at-risk in the community.

The team administering the Ministry of Justice's (MoJ's) PNC research database created a set of data files containing cases where the offender's conventional reoffending follow-up would have started between March 2005 and March 2008. These offenders either were cautioned or sentenced to a non-custodial disposal between these dates, or were discharged from custody between these dates. Data were drawn from the PNC between 28 and 30 March 2011; follow-ups commencing in March 2008 therefore had a 'buffer period' of a full year, during which offences committed in the follow-up period could result in a PNC-recorded conviction. An index date – the date of non-custodial disposal or discharge from custody, which would therefore be the start date for a conventional first-two-year reoffending study – was identified for each case. At this point, offenders were included on multiple occasions if they had more than one index date.

This report constructs predictors of reoffending which use '***next-two-year***' follow-ups. These predictors estimate the offender's likelihood of reoffending specifically for the two years following the point they have reached in the post-sentence or post-discharge process. The falling hazards of reoffending as offence-free time increases after sentence or discharge are incorporated into the predictors' scoring system. As the maximum lengths of both Youth Rehabilitation Orders and (adult) Community Orders are three years, cases were included in the next-two-year follow-up sample when their follow-up start date was no more than three years after their date of sentence/discharge.[113]

Offenders who were cautioned/sentenced/discharged prior to March 2008 were traced until the equivalent date in March 2008; they were retained in the sample for the 'next-two-year' analysis

[112] Potentially, the predictors could also be used in relation to non-conviction sanctions. "Diversion of lower-risk offenders from courts and probation" (House of Commons Justice Committee, 2011, para. 164) has been promoted as a desirable goal, but requires valid identification of such offenders to avoid damaging public safety and confidence in the justice system.

[113] A small number of former prisoners are supervised (or at least required to maintain contact with criminal justice agencies, if not actively supervised) for more than three years after discharge, having received determinate sentences of at least six years, indeterminate sentences for public protection or life sentences. These offenders will be assigned a next-two-year reoffending estimate equal to that of those discharged exactly three years previously.

provided they did not commit a proven reoffence of any type,[114] nor were imprisoned for any offence,[115] prior to that date. Each offender's ***offence-free time*** was the period between their sentence/discharge and their follow-up start date, and the method used here ensured that its duration was always an exact number of months. For example, offenders sentenced/discharged on 16 June 2006 were traced until 16 March 2008. If they neither committed a new offence (whether proven before or after 16 March 2008) nor were imprisoned, then they were included in the next-two-year follow-up with a follow-up start date of 16 March 2008 and an offence-free time of 21 months. Offenders remained in the sample for the subsequent sanction provided that they were resanctioned/discharged prior to March 2008.[116] The remaining set of 1,809,301 offenders, with one case each, therefore formed an appropriate dataset for next-two-year reoffending calculations.[117] Random numbers were used to assign these cases to construction (two-thirds) and validation (one-thirds) sub-groups, providing 1,205,340 construction and 603,961 validation cases.

Among all final cases, 22% of offenders were female. Mean age at first sanction was 22.7, at current sanction 28.4 and at index date 28.5. Current offence categories are shown in Table 8.4 below. The current disposal was a caution, reprimand or final warning in 45% of cases and a conviction in the remaining 55%. The sentence breakdown for convicted offenders was as follows:

- 15% discharged from court;
- 37% fined;
- 20% given an adult community sentence;
- 8% given a youth community sentence or referred to a youth offender panel;
- 5% given a suspended sentence;
- 11% sentenced to immediate custody; and
- 3% otherwise sentenced.

Procedure

Previous sanctions and proven reoffending

In the next-two-year follow-up, proven reoffending comprised offences committed between each offender's follow-up start date and the same date two years later. General and violent next-two-year proven reoffending statuses were both calculated. In this chapter, 'violent' reoffending always refers to proven OVP-class offending, adopting the recommendation set out in Chapter 5 that arson offences should be included. The numbers of general and violent sanctions in each offender's career were

[114] When any type of proven reoffending is recorded, predictors of further reoffending should be rescored. This can therefore be considered a new follow-up period, even when the reoffending results only in the continuance of existing sentences.

[115] It is possible for an offender to be imprisoned without committing a new offence, as the result of a pseudoreconviction. Pseudoreconvictions are offences committed before the start of the period of interest but brought to justice after. Information on periods on remand in custody, which could also disrupt a reoffending follow-up, was not available.

[116] However, an individual who committed a reoffence before their follow-up start date for which they were sanctioned/discharged after March 2008 would not be represented in the next-two-year dataset at all.

[117] A small number of cases were removed following data consistency checks (e.g. offence date is realistic in relation to offender's date of birth).

counted, as was the number of cautions. Age at first sanction, current sanction, and at the index date were calculated.

Checking the predictive value of offence-free time

A simple exercise was undertaken to establish the value of the offence-free time item, using the construction sample. Two models were fitted for any proven reoffending, one using only OGRS3 scores as a predictor, and the other using OGRS3, offence-free months and the square of offence-free months as predictors. A similar pair of models was fitted for proven violent reoffending. The residuals – the differences between actual and predicted reoffending in each month – were examined. If offence-free time is a worthwhile risk factor, then the models containing only OGRS3 scores would produce clinically and statistically significant residuals while those also involving the month count would have very small residuals.

In addition to these initial tests, versions of the final OGRS4/G and OGRS4/V models were fitted without the offence-free time component. This serves two purposes: an additional check on the value of offence-free time, and also for use in the development of version 2 of OGP and OVP (see Chapter 9).

Selection of static and offence-free time risk factors, and exploratory data analysis

In this chapter, ***primary risk factors*** are those which are coded and entered by the assessor, and ***secondary risk factors*** are those which are calculated through IT functionality (whether this is in a spreadsheet or an operational IT application). The choice of primary static risk factors was constrained to ensure that all could be coded swiftly by practitioners or administrative staff on the basis of summary printouts of individual offenders' demographics and criminal histories. Such printouts are routinely available from the operational PNC.

A much wider range of secondary criminal career, offence-free time and age risk factors was produced from the small number of primary risk factors. As the computation of OGRS scores is fully automated in routine IT systems, these secondary risk factors can be mathematically complex and need not be amenable to manual scoring. However, while many predictive methods proposed in recent years effectively reduce the scoring process to a 'black box' (e.g. the neural networks used by Yang, Liu and Coid, 2010), it was viewed as preferable that the OGRS4 models should make sense to practitioners when their mathematical intricacies are explained or simplified.

The caution count was used to calculate a 'current caution' (rather than 'current conviction') item, and indicator variables created for each combination of cautions and convictions for those with no more than three total sanctions. Indicator variables were created for first- to seventh-time entrants for any offence, and for first- to fifth-time entrants for violent offences. The number of years between the first and current sanctions was calculated in its own right, and used as the basis for the OGRS3 Copas rate (see below) and numerous variants. A measure of violent specialisation – the proportion of all sanctions which involved violent offences – was calculated.

The ***Copas rate*** is a term used to describe various mathematical functions which have been developed to summarise the volume and speed of an offender's known criminal career. A different version has been used in each version of OGRS to date.[118] In each version, offenders with more criminal previous sanctions have a higher 'rate', as do those whose criminal career spans a short number of years. The ***criminal career length*** is measured as the difference between age at first sanction to age at current sanction, so does not lengthen while the custodial portion of a current sentence is being served. The third age factor used is ***age at index date***, which is the date of the offender's non-custodial sentence or discharge from custody.[119] Whereas age at assessment date would be time-dependent, but would confound the effects of age with those of offence-free time (because it would equal age at index date plus offence-free time), age at index date is a static factor as it will not change before reoffending occurs. The use of age at index date is consistent with OGRS3.[120]

Initial models were run with offences categorised in the same way as in OGRS3. Residuals were then examined for individual offences within each OGRS3 category. Where the initial models under- or over-predicted both general and violent reoffending, to a statistically significant extent, the creation of new offence categories and the movement of offences between similar categories were considered. Checks were made of the viability of both the proposed categories and the remnant categories from which offences had been removed, ensuring that each new category should include at least one percent of offenders, and remnant categories should do so unless they dealt with offences likely to cause serious harm and were already below one percent in OGRS3.

Offence-free months was calculated as the number of months between the index date and follow-up start date. This was then squared, cubed, and raised to the fourth power. Age was treated as a continuous variable, and also squared, cubed and raised to the fourth power. In past versions of OGRS, age has been grouped – for example, OGRS3 has eleven age groups, ranging from 10–11 to 51+. Practitioner feedback suggested that this was viewed as having perverse effects in some cases, with the scores of repeat offenders suddenly reducing when a new offence led to their score being based on an older age group, or scores being highly dependent on the exact calculation of a custodial release date. Treating age as continuous ensures that single-year age differences only have a small effect on predicted probabilities.

---

[118] The OGRS3 Copas rate formula is log((sanction count/(career length + 10)). The constant was set to all values between 1 and 40 [a zero value would result in division by zero for some offenders], while the logarithm was replaced with a square root and with no transformation. Versions were also created for violent sanctions.

[119] In practice situations where the release date of an imprisoned offender is not yet known, OGRS3 user guidance states that the index date should be set to the earliest plausible release date. OGRS4 will follow this convention.

[120] In OGRS, all ages are always coded as age at last birthday, and as such all criminal career lengths are integers.

Scoring the extent and speed of criminal careers

In OGRS3, the only criminal career items included are the Copas rate, current offence category and four indicator variables relating to current cautions and early career. If applied to all offenders, the selection of the Copas rate has to juggle several requirements. The choice of transformation (log, square root or none) and constant in the denominator must together correctly baseline the likelihood of reoffending of first-time entrants (where no further relevant information is available), second-time entrants (with variance in the number of years between the first and current sanction) and repeat offenders (those with three or more sanctions, with variance in both the sanction count and the first-to-current-sanction year count). It may be the case that the likelihoods for first- and second-time entrants are not optimised by using a transformation and constant which are largely determined by the more numerous (and more frequently recidivist) repeat offenders.

Therefore, an alternate approach was trialed. Indicator variables were used to identify first- and second-time entrants and repeat offenders. The models included the first-time indicator, an interaction between the second-time indicator and the number of years between first and current sanction, and an interaction between the repeat offender indicator and the Copas rate. Therefore, the Copas rate was only applied to repeat offenders, for whom it was genuinely meaningful, while early-career (not necessarily young-age) offenders were separated in this part of the model.[121]

It was also recognised that the effect of criminal history might be different for female offenders. Therefore, interactions between the above model terms and gender were added to the model. Checks on model residuals were carried out, and model specifications were changed and the model rerun to deal with any identified problems with absolute predictive validity.

Model selection

Models were selected through backward stepwise regression, with p=.05. In practice, numerous decisions needed to be taken around the choice of model terms, and the Akaike Information Criterion (AIC) statistic was used to differentiate between alternate models where all terms in the competing models were statistically significant.[122] An element of expert judgement was also applied, with preference being given to simpler models when models appeared to be overfitted (i.e. overly complex and at risk of not predicting well for new data).

Consistency of predictions of general and violent proven reoffending

As the OGRS4/G and OGRS4/V predictive models were generated independently from one another, it was possible that some offenders could receive an OGRS4/V prediction higher than their OGRS4/G

[121] At an early stage in the modelling process, entirely separate static models were developed for first-time and second-time offenders. The resulting improvement in model fit was not considered great enough to justify the increase in explanatory complexity, e.g. "Offence A is associated with greater risk among first-time offenders and lower risk among second-time offenders, while Offence B shows the opposite pattern and Offence C is associated with higher risk for all but second-time offenders."

[122] AIC is a measure of the relative goodness of fit of a statistical model, considering the complexity and accuracy of the model.

prediction. Given that each offender's true probability of violent reoffending cannot be higher than their true probability of any reoffending, it was recognised that this would be unsatisfactory. Checks were therefore made to determine the proportion of cases in which this anomaly occurred. The practical remedy in this situation was to increase the OGRS4/G score to equal the OGRS4/V score. The comparisons in 'Model validation' below use these corrected OGRS4/G scores, and the correction should be considered an integral part of the OGRS4/G scoring algorithm.

Model validation

To test relative risk, Area Under Curve (AUC) statistics were used. AUCs are very dependent on the true distribution of risk in the sample being studied, and therefore this study tested whether predictive validity had improved by directly comparing old and new predictors on the validation sample.[123] OGRS4/G was compared with OGRS3, while OGRS4/V was compared with OGRS3, RM2000/v and the static subscale of OVP1.[124] As the predictive scores were correlated, all significance tests for differences in AUCs were conducted using a non-parametric comparison developed by DeLong, DeLong and Clarke-Pearson (1988). To indicate whether improvements in predictive validity were due to the offence-free months term or other changes to the predictors, AUCs were also produced for logistic regression models which combined existing predictors with offence-free month polynomials.[125]

To validate prediction of absolute risk, mean predicted and observed proven reoffending rates were compared for age/gender sub-groups. Small residuals indicate good model fit. Relative predictive validity within each sub-group was estimated by calculating the AUC for cases in that sub-group. To test whether relative predictive validity was higher in one age/gender group than another, T-tests were used (Gönen, 2007), comparing the AUC of each group against that of the 21–24 age group for the same gender.

Additional analysis is presented in the form suggested by the Hosmer-Lemeshow test (Hosmer and Lemeshow, 2000). Predicted probabilities were ranked and placed into deciles (ten equal-sized groups). A chi-square test is usually then run to compare actual and predicted rates, determining whether predictions prove to be accurate across the risk range. While this is redundant when analysing huge samples (as statistically significant differences are inevitable even when the model fits very well in practical terms), scrutiny of the rates for each decile remains of value.

[123] In the validation sample for OGRS3 (Howard *et al*., 2009), it had an AUC of 0.80 as a predictor of all proven reoffending. In the validation sample for the static/dynamic OASys-based predictors OGP1 and OVP1 (Howard, 2009), OGP1 had an AUC of 0.80 as a predictor of proven non-violent reoffending, and OVP1 had an AUC of 0.74 as a predictor of proven violent reoffending.

[124] These predictors were selected because, like OGRS4, they are scored using only static risk factors and therefore are operationally viable in the same situations as OGRS4. Additionally, these scores can be exactly simulated from PNC data. Other predictors, such as OVP1's full static/dynamic score, lack both of these qualities.

[125] As these models are not operationally available, they were not included in the DeLong *et al.* (1988) comparisons.

8.3 Results

Checking the predictive value of offence-free time

Table 8.2 summarises the simple logistic regression models which were fitted (inputting OGRS3 scores into the models) to provide an initial check that accounting for offence-free months improved prediction.[126] Exploratory data analysis showed that modelling offence-free time over its 36-month range was more successful when either only the simple term and the squared term were used, or all four terms were used.[127] The models indicate that offence-free time makes a non-negligible improvement to the prediction of both outcomes.[128]

Table 8.2: Effect of supplementing OGRS3 with offence-free time when predicting all and violent reoffending

Type of reoffending predicted	Model parameter and logistic regression coefficient				AUC
	Intercept	OGRS3	Offence-free months	Offence-free months, squared	
Any (no offence-free terms)	-2.7631	0.0409	n/a	n/a	0.750
Any (with offence-free terms)	-2.1309	0.0391	-0.0525	0.000644	0.758
Violent (no offence-free terms)	-3.2287	0.0359	n/a	n/a	0.733
Violent (with offence-free terms)	-2.6430	0.0337	-0.0462	0.000513	0.741

Figure 8.1 compares actual and predicted general and violent reoffending rates for cases with different lengths of offence-free time. When offence-free time was not included in the predictive model, it failed to fully capture the reduction in reoffending: for general reoffending, the predicted rates fell from 31% to 21% between zero and 36 offence-free months,[129] while the actual reoffending rate fell from 43% to 14%. By contrast, the model which included offence-free time did keep pace with the actual fall, with these predicted rates falling from 41% to 14%. Similar results were found for violent reoffending. For both types of reoffending, the model incorporating offence-free time was able to track the true reoffending rate closely.

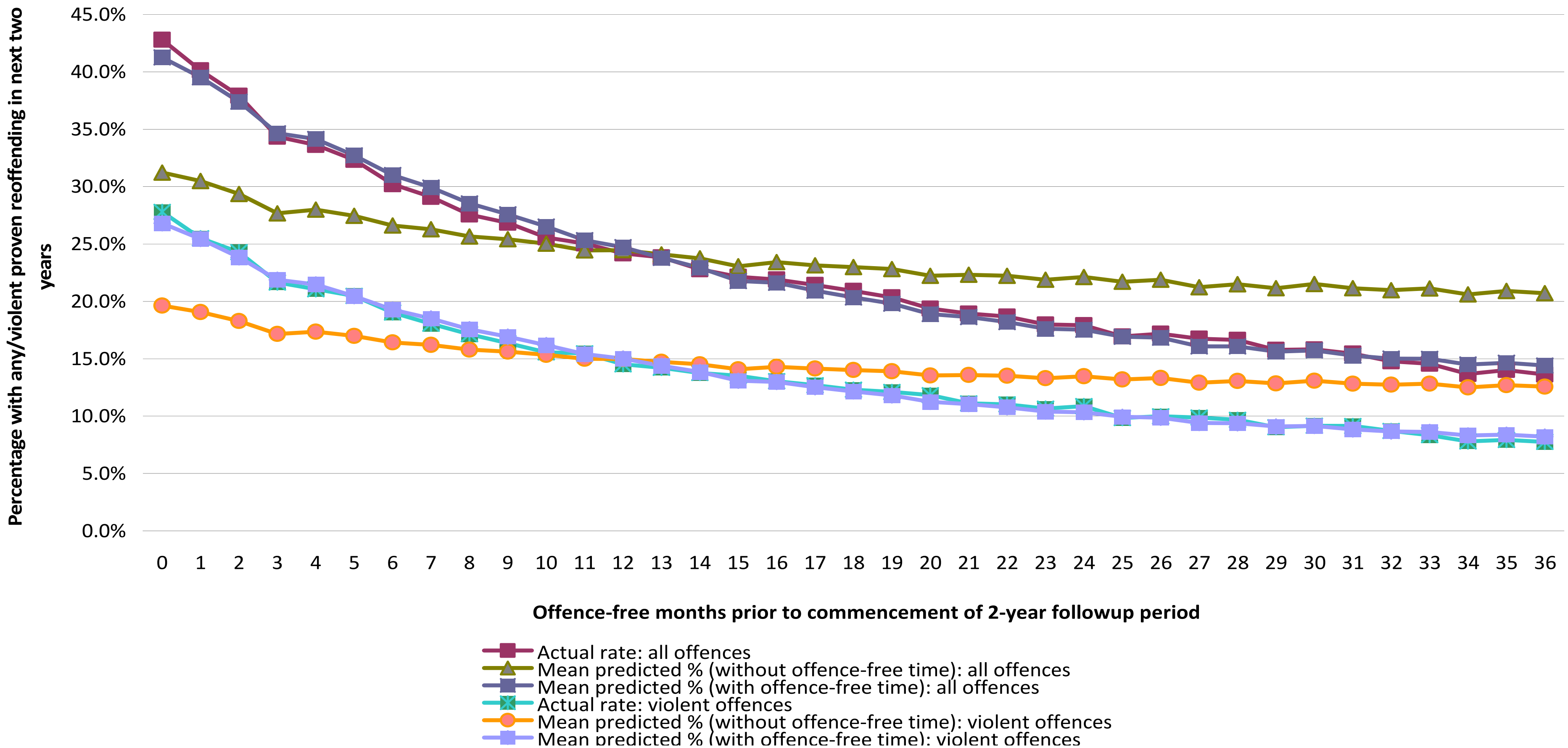
126 For example, the probability of reoffending in the "any (with offence-free terms)" model was $e^z/(1+e^z)$, where z = -2.1309 + (0.0391 * OGRS3 2-year % score) + (-0.0525 * offence-free months) + (0.000664 * (offence-free months, squared)).

127 The intermediate option – using the simple, squared and cubed terms – produced algorithms where the modelled probability of reoffending would increase at the highest number of offence-free months, whereas the two- and four-term algorithms produced steady declines as offence-free months increased.

128 It is notable that the AUCs for the prediction of any proven reoffending are well below those in the OGRS3 report. This suggests that the current sample is more homogeneous, making it more difficult to separate likely reoffenders and non-reoffenders using any given risk predictor.

129 Predicted rates still fell because those with higher OGRS3 scores were more likely to reoffend early and therefore not be in the dataset with higher values of offence-free time.

Figure 8.1: Actual 2-year proven reoffending rates, and predicted rates based on OGRS3 with and without offence-free months



Selection of static and offence-free time risk factors, and exploratory data analysis

Table 8.3 presents the general and violent reoffending rates associated with a range of static risk factors. These results are illustrative, as they do not take into account variations in other risk factors. The effects of offence-free time were filtered out to facilitate comparison, with only March 2008 starts (with zero offence-free months) included (N = 74,247).

Table 8.3: General and violent two-year reoffending rates by static risk factors (offenders with zero offence-free months)

Risk factor	Characteristic	% with this characteristic	Proven reoffending	
			Any offences	Violent offences
(None)	***All offenders***	***100%***	***42.8%***	***27.7%***
Gender	Female	19%	30.7%	18.5%
	Male	81%	45.6%	29.9%
Current sanction	Caution	40%	29.7%	19.4%
	Conviction	60%	51.6%	33.4%
Offence category	Absconding/bail	2%	66.0%	43.3%
	Acquisitive violence *	1%	57.9%	39.9%
	Burglary (domestic)	1%	67.7%	37.1%
	Burglary (other)	2%	63.7%	36.6%
	Criminal damage *	10%	45.8%	35.6%
	Drink driving	7%	21.1%	11.1%
	Drug import/export/production	1%	37.3%	17.9%
	Drug possession/supply	10%	44.3%	22.2%
	Drunkenness *	3%	57.9%	50.0%
	Fraud, forgery and misrepresentation	3%	24.7%	11.0%
	Handling stolen goods	1%	56.4%	33.3%
	Motoring (not drink driving)	4%	39.2%	19.2%
	Other	2%	44.8%	30.1%
	Public order and harassment *	9%	45.7%	36.5%
	Sexual (against child)	0.4%	20.8%	8.3%
	Sexual (not against child)	0.4%	25.9%	12.4%
	Theft	15%	54.4%	28.7%
	Vehicle-related theft	3%	61.2%	39.3%
	Violence against the person *	24%	34.6%	26.2%
	Welfare fraud	1%	8.0%	4.3%
Total sanctions	1	33%	19.9%	12.0%
	2	15%	34.1%	21.8%
	3	9%	43.5%	28.9%
	4 or 5	11%	50.6%	34.4%
	6 to 9	12%	59.4%	40.5%
	10 to 19	12%	67.9%	44.3%
	20 to 29	4%	76.7%	45.9%
	30 and over	3%	87.7%	57.1%

Risk factor	Characteristic	% with this characteristic	Proven reoffending	
			Any offences	Violent offences
Violent sanctions	0	25%	25.0%	10.5%
	1	29%	31.0%	19.2%
	2	14%	47.0%	30.7%
	3	9%	55.8%	38.6%
	4 or 5	10%	62.9%	43.0%
	6 to 9	8%	70.9%	52.6%
	10 to 19	4%	78.5%	63.8%
	20 and over	1%	94.3%	87.9%

Note. Offence categories marked * are considered violent.

Of the OGRS3 offence categories, one (soliciting/prostitution) was abolished entirely. Two were merged into a single OGRS4 category (vehicle-related theft), while two new categories were created (welfare fraud and drunkenness). Consequently, as shown by Table 8.4, 20 offence categories remained. The recategorisation led to about seven percent of offenders substantively changing category.[130] All of the new categories contained more than one percent of the total OGRS4 sample. Table 8.3 shows that reoffending rates were low among those sanctioned for sexual offences, drink driving and fraud-related offences, and lowest of all among welfare fraud cases. The separation of welfare fraud from other fraud cases will improve prediction most for female offenders, as welfare fraud was the primary offence for under one percent of males but three percent of females. By contrast, over 90% of cases of sexual offending (both groups), non-domestic burglary, vehicle theft and non-drink driving motoring involved male offenders. Overall reoffending rates were highest among those sanctioned for absconding, burglary (both groups) and vehicle-related theft. Those sanctioned for acquisitive violence and drunkenness had slightly lower general reoffending rates, but were among the most frequent violent reoffenders. Specialisation in violent offending was also evidenced by the lower ratio of general to violent reoffending among the criminal damage, public order and violence against the person groups, among whom this ratio was around 4:3 compared with around 2:1 for most motoring, acquisitive and sexual offence groups. This ratio was especially low among those sanctioned for drunkenness. Checks confirmed that these offenders did not specialise particularly in drink-related offences: they also had high rates of an alternate violent reoffence status, which excluded offences from the drunkenness group.

[130] This seven percent does not include offenders in the present two motor theft categories moving to the new single vehicle-related theft category, nor relabelling without moving substantive category (e.g. moving existing public order offenders to 'public order and harassment').

Table 8.3 also shows that reoffending was less frequent among female offenders and those cautioned. Offenders with longer sanction histories had higher general and violent reoffending rates. Specialisation in violence is again clear: the majority of reoffenders with no previous violent sanctions did not reoffend violently, whereas at least four-fifths of reoffenders with ten or more previous violent sanctions were violent again. The distribution of sanction histories is notable: about half of all offenders had only one or two total sanctions, and half had no history of violence or one violent sanction, but long tails existed with many offenders exceeding 20 total sanctions or 10 violent sanctions.

For general reoffending, Figure 8.2 offers further detail on the association between criminal career and recidivism. Offenders are separated by both the calendar length of their criminal career and the number of sanctions within that career, with only combinations represented by at least 50 offenders being displayed. Among those whose sanctions fell within a single year, reoffending rates ranged from 20% for first-time entrants to 73% for those with four or five sanctions, and similarly large differences can be found at higher sanction counts. The graph can also be read crossways, holding the number of sanctions constant:

- Those with two sanctions in a single year had a 45% reoffending rate, whereas those with two sanctions separated by at least 20 years had just a 7% rate.
- For three sanctions, rates of 67% and 22% respectively were found.
- Negative associations between career length and recidivism could be observed even at much higher sanction counts.

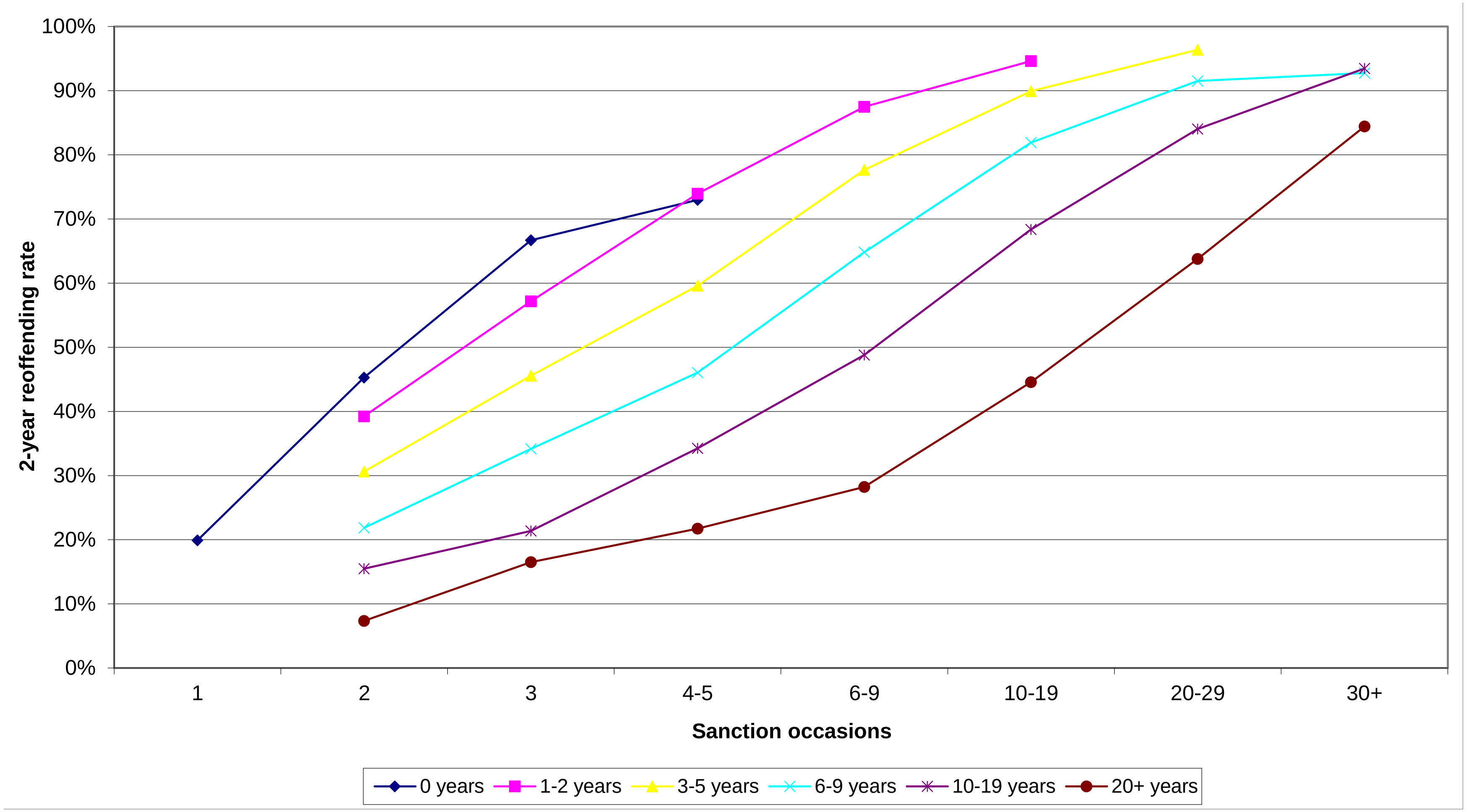
In other words, those with a known history of offending at a rapid tempo (every few months, rather than every few years) were more likely to reoffend at least once within the next two years.

Table 8.4: OGRS3 and OGRS4 offence categories

OGRS3 category	Closest OGRS4 category	Changes to category composition		% of offenders in category	
		Offences removed	**Offences added**	**OGRS3**	**OGRS4**
Absconding/bail		*None*	Refusing/failing to give drug test sample	1.6%	1.7%
Burglary (domestic)		*None*	*None*	1.0%	1.0%
Burglary (other)		*None*	*None*	1.2%	1.2%
Drink driving		*None*	*None*	10.9%	10.9%
Drug import/export/production		*None*	*None*	0.9%	0.9%
Drug possession/supply		Refusing/failing to give drug test sample	*None*	8.1%	8.0%
Fraud and forgery	Fraud, forgery and misrepresentation	Dishonest representation for purpose of obtaining benefit	Perjury; Immigration Act offences; theft by an employee; money laundering; frauds relating to vehicle registration, licence etc.	3.1%	3.8%
Handling stolen goods		*None*	*None*	1.1%	1.1%
Motoring (not drink driving)		Interference with motor vehicle; being carried in vehicle taken without consent; frauds relating to vehicle registration, licence etc.	*None*	5.2%	4.6%
Other offences		Perjury; Immigration Act offences; Social Security fraud; Improper use of public electronic communications network; failure to comply with Football Banning Order; summary racially aggravated harassment; malicious communications; use of violence to gain entry	Soliciting/prostitution offences	2.8%	2.2%
Public order	Public order and harassment	Drunkenness (simple); drunk and disorderly	Harassment; intentional harassment/alarm/distress; breaches of relevant orders; putting people in fear of violence; similar racially aggravated offences; improper use of public electronic communications network; failure to comply with Football Banning Order; summary racially aggravated harassment; malicious communications	9.6%	9.0%

OGRS3 category	Closest OGRS4 category	Changes to category composition		% of offenders in category	
		Offences removed	**Offences added**	**OGRS3**	**OGRS4**
Robbery	Acquisitive violence	*None*	*None: the name change reflects the longstanding inclusion of aggravated burglary*	0.7%	0.7%
Sexual (against child)		*None*	*None*	0.5%	0.5%
Sexual (not against child)		*None*	*None*	0.6%	0.6%
Soliciting/ prostitution	(None: category removed)	All such offences	*None*	0.3%	n/a
Theft		Theft of bicycle; theft by an employee; money laundering	*None*	13.7%	12.3%
Theft of vehicle AND theft from vehicle	Vehicle-related theft	*None: all offences from the two old categories are included in the single new category*	Theft of bicycle; interference with motor vehicle; being carried in vehicle taken without consent	1.2%, 0.5%	2.2%
Violence against the person	Violence against the person (not harassment)	Harassment; intentional harassment/alarm/ distress; breaches of relevant orders; putting people in fear of violence; similar racially aggravated offences	*None*	27.6%	26.3%
Criminal damage		*None*	Use of violence to gain entry	9.6%	9.7%
(None: new category)	Drunkenness	*None*	Drunkenness (simple); drunk and disorderly	n/a	2.1%
(None: new category)	Welfare fraud	*None*	Dishonest representation for purpose of obtaining benefit; Social Security fraud	n/a	1.2%

Figure 8.2: Two-year general reoffending rate by number of sanctions and length of criminal career



Scoring the extent and speed of criminal careers

The final OGRS4/G model selected included several risk factors related to criminal careers. Currently cautioned offenders had somewhat lower reoffending probabilities than currently convicted offenders, controlling for all other factors. The reworking of the model for first- and second-time and repeat entrants was successful, with the indicator variables for first- and second-time entrants being significant, as was the interaction between second-time entrant status and years between first and second sanction. Indicator variables for third-time and higher order statuses were not significant or were removed due to non-helpful parameters. A variant of the Copas rate was included (see note below Table 8.5). In addition to these terms, the number of sanctions (as a simple term, without any mathematical transformation)[131] was a significant predictor with a positive regression parameter. This suggests that the Copas rate's rendering of the complex relationship between sanction count and duration of criminal career does not fully reflect the importance of sanction count.

Two interactions between gender and criminal career were also included in the selected OGRS4/G model. The second-time/years since first sanction interaction and Copas rate both had terms which meant that they was less predictive for female offenders.

The handling of age proved difficult. It proved impossible to fit a single set of polynomial terms which reflected both (i) the low level of age-related variation among juveniles and (ii) the sharp decrease in reoffending rates with increasing age in young adulthood. Therefore, separate terms were fitted for under-18s and those aged 18 and over, with additional indicator variables to cover ages 10 and 11. Most of these terms were also fitted separately for male and female offenders, but the age 10 and 11 indicators were fitted once only due to the small number of female offenders at these ages.

The OGRS4/V model included all the OGRS4/G model terms, though the chosen Copas rate used a different formula to the OGRS4/G Copas rate. In addition, never-violent and once-violent indicator variables were significant, as was violent sanction count. A violence-specific Copas rate variant was also significant. This kept the usual definition of criminal career length as the basis of the denominator, but the violent sanction count replaced the total sanction count as the numerator. It was fitted for all offenders (i.e. not just those with 3+ violent sanctions), as exactly matching the set-up of the general criminal career items would require the collection of new primary risk factors. One gender interaction proved to be significant: never-violent women were particularly unlikely to reoffend violently.

[131] The square of sanction count was also included, but failed to improve model AIC.

Model selection

Table 8.5 details the finally selected models. Odds ratios are provided for most items, but not for age and offence-free time because the odds ratio for an increase of one year or offence-free month is dependent on the starting age/time value. Likewise, an increase of one general or violent sanction affects both the simple sanction count and the Copas rate(s), interacting with criminal career length in the latter instance. Odds ratios are also not quoted for gender due to its interaction with age and Copas rate.

Key differences from the OGRS3 model (Howard *et al.*, 2009) are as follows:

- the inclusion of offence-free time;
- the inclusion of sanction counts and indicator variables as well as the Copas rates; and
- the modelling of age through continuous polynomials rather than through discrete groups.[132]

The differences between offence parameters are smaller in OGRS4/G than in OGRS3 if the very low parameters of the new welfare fraud category are discounted, most likely because of the greater range of other risk factors taken into account by OGRS4. They are also smaller in OGRS4/V than OGRS4/G; they may have been partially displaced by the violent offence history items, which appear in ‘V’ but not ‘G’.

Table 8.5: OGRS4/G and OGRS4/V logistic regression model parameters

Risk factor	Value/unit	Outcome					
		General			Violent		
		B	***SE***	***Exp (B)***	***B***	***SE***	***Exp (B)***
Intercept		7.98337	0.0818		7.54632	0.195	
Age at index date	10	-0.1441	0.0607		-0.13339	0.0672	
	11	-.01305	0.0382		-0.11527	0.0425	
Male, aged <18		5.47899	0.5011		-4.5032	0.5599	
Age at index, if	Simple	0.17949	0.0649		-0.064393	0.0721	
male aged <18	Squared	-0.00932605	0.0022		-0.0127746	0.00244	
Female, aged <18		0.26869	0.7154		-1.06687	0.8328	
Age at index, if	Simple	-0.67795	0.097		-0.52297	0.113	
female aged <18	Squared	0.017875	0.00335		0.010999	0.00391	
Age at index date,	Simple	-0.53236	0.0176		-0.6317	0.0208	
if male aged 18+	Squared	0.016939	0.00068		0.020864	0.000805	
	Cubed	-0.000238996	0.000011		-0.000298183	0.000013	
	To power 4	0.000001194	6.28E-08		0.000001492	7.46E-08	

[132] As an example, the formulas in Howard *et al*. (2009) for OGRS3, and this report for OGRS4/G, were applied to a male offender with zero offence-free months, age at first sanction of 13, eight previous sanctions and a current public order offence. Under OGRS3, with ages at current sanction and index date of 17 his two-year predicted probability would be 87%, while with those ages raised to 18 this probability would be 78%. Under OGRS4/G, these would be 84% and 80% respectively. Under OGRS3, with both ages at 39 his probability would be 39% and at age 40 it would be 35%. Under OGRS4/G, these would be 45% and 44% respectively. Therefore, at these age category boundaries, one-year age increases reduced predictions by 9% and 4% using OGRS3, compared with just 4% and 1% respectively using OGRS4/G.

Risk factor	Value/unit	Outcome					
		General			Violent		
		B	***SE***	***Exp (B)***	***B***	***SE***	***Exp (B)***
Female, aged 18+		-5.69369	0.4376		-4.55932	0.538	
Age at index,	Simple	-0.05792	0.0455		-0.26288	0.0563	
if female aged	Squared	0.002453641	0.00179		0.010233	0.00221	
18+	Cubed	-0.000050561	0.00003		-0.000162944	0.000036	
	To power 4	0.000000304	1.73E-07		0.000000869	2.12E-07	
Offence category	Absconding/bail	0.26622	0.0155	1.31	0.29589	0.0177	1.34
	Acquisitive violence	0.25561	0.0235	1.29	0.11664	0.0258	1.12
	Burglary (domestic)	0.28426	0.0193	1.33	0.1756	0.022	1.19
	Burglary (other)	0.23317	0.0178	1.26	0.18374	0.0206	1.20
	Criminal damage	0.11569	0.00835	1.12	0.16348	0.0104	1.18
	Drink driving	-0.20591	0.00967	0.81	-0.11645	0.0127	0.89
	Drug imp./exp./product.	-0.1107	0.0236	0.90	-0.26348	0.0322	0.77
	Drug possession/supply	0.09829	0.00883	1.10	-0.0633	0.0114	0.94
	Drunkenness	0.21106	0.0154	1.23	0.39985	0.017	1.49
	Fraud, forgery and misrepresentation	-0.26415	0.0144	0.77	-0.27336	0.0199	0.76
	Handling stolen goods	0.11709	0.0206	1.12	0.12049	0.0248	1.13
	Motoring (not drink driving)	-0.03723	0.0112	0.96	-0.07403	0.0146	0.93
	Other	-0.07144	0.0165	0.93	-0.00175	0.0202	1.00
	Public order, harassment	0.08426	0.00852	1.09	0.19568	0.0105	1.22
	Sexual (not against children)	-0.16052	0.0375	0.85	-0.2128	0.0494	0.81
	Sexual (against children)	-0.33557	0.0374	0.71	-0.60348	0.0566	0.55
	Theft	0.13227	0.00797	1.14	0.09868	0.0102	1.10
	Violence against the person	0.02911	0.00651	1.03	0.10912	0.00868	1.12
	Vehicle-related theft	0.19396	0.014	1.21	0.22343	0.0162	1.25
	Welfare fraud	-0.83547	NA		-0.47395	NA	
Current sanction is...	Caution	-0.16361	0.00624	0.85	-0.05881	0.00743	0.94
Sanction count	1st	-3.37339	0.0257	0.03	-1.68110	0.0278	0.19
	2nd	-2.62861	0.0251	0.07	-1.08040	0.0261	0.34
	Per sanction	-0.0082958	0.00083	0.99	-0.014046	0.00122	0.99
Years since... (2nd only), if female	Per year	-0.012296	0.00302	0.99	-0.00827	0.00425	0.99
Years between 1st and index sanction (2nd sanction only), if male	Per year	-0.026684	0.00161	0.97	-0.01807	-0.00202	0.98
Tempo (see Note) (3+ sanctions only), if female	Per unit of general Copas rate	0.95735	0.0127	2.60	0.44664	0.0122	1.56
Tempo, if male	Per unit of gen. Copas	1.03816	0.0112	2.82	0.49063	0.0152	1.63
Offence-free months	Simple	-0.093578	0.0034		-0.075169	0.00387	
	Squared	0.00485345	0.000427		0.003244689	0.000496	
	Cubed	-0.000141074	0.000019		-0.000078397	0.000022	
	To power 4	0.000001479	2.76E-07		0.000000648	3.27E-07	
Violent sanction count	None, if female	n/a	n/a		-1.80039	0.0576	0.17
	None, if male	n/a	n/a		-1.42630	0.0555	0.24
	One	n/a	n/a		0.071643	0.0148	1.07
	Per sanction	n/a	n/a		0.042926	0.00288	1.04
Violent sanction tempo (see Note)	None	n/a	n/a		-0.37409	0.02	1.43

Note. Copas rate = log ((count of all sanctions or violent sanctions)/(age at current sanction + *x* - age at first sanction)). Log is the natural logarithm (i.e. to base *e*, not base 10). In OGRS4/G, *x*=26. In OGRS4/V, *x*=12 for the general Copas rate, *x*=30 for the violent Copas rate.

Consistency of predictions of general and violent proven reoffending

OGRS4/V predicted probabilities were higher than OGRS4/G predicted probabilities for 0.17% of offenders. In half of these cases, OGRS4/V was one percent higher than OGRS4/G, and in half of the remainder it was two or three percent higher. The subsequent adjustment which raises OGRS4/G scores to equal OGRS4/V scores is therefore used very rarely and has little impact on predictive validity metrics, although it is vital to avoid confusion in those cases where it is required.

Model validation

Table 8.6 presents Area Under Curve (AUC) statistics for the prediction of general and violent reoffending. For general reoffending, OGRS4/G clearly improves upon OGRS3 among all cases. The intermediate AUC of the OGRS3 and offence-free month hybrid model illustrates that OGRS4/G brings benefits from both the offence-free month component and the rest of the model structure. However, these general modelling benefits may be most relevant to offenders sanctioned some time ago, as OGRS4/G only brings a marginal benefit (.002 points of AUC) for those with zero offence-free months. This may be because OGRS3 was optimised for use with offenders with zero offence-free months, whereas OGRS4 was constructed from cases with differing offence-free month periods.

For all violent reoffending, a wider range of feasible predictors are available. The static scale of OVP1 was considered to be the strongest existing static predictor, and proved to be so, with an AUC significantly greater than those of OGRS3 or the 0–8 RM2000/v score. Both OGRS4/G and OGRS4/V were more predictive than static OVP1. The intermediate result for the static OVP1 / offence-free month combination suggests that OGRS4/V's improvement was due to both offence-free months and improved coding of static factors.

Further DeLong *et al.* (1988) tests compared the two OGRS4 predictors for predicting all violent reoffending, confirming that OGRS4/V was more predictive than OGRS4/G among (i) all cases and (ii) the month zero cases only (both p<.0001). The difference between OGRS4/V and OGRS4/G was greater among month zero cases (0.020 points of AUC) than all cases (0.012 points).[133]

Table 8.6: Relative predictive validity: Area Under Curve (AUC) statistics for all and violent proven reoffending

Predictor	General reoffending		Violent reoffending	
	All cases	Month zero cases only	All cases	Month zero cases only
General reoffending predictors				
OGRS3	0.750 (R)	0.795 (R)	0.724 ***	0.749 ***
OGRS3 and OFM	0.759 (NC)	n/a	0.733 (NC)	n/a
OGRS4/G	0.770 ***	0.797 **	0.757 ***	0.759 **

[133] OGRS4/V was also found to be the best predictor of homicide/wounding reoffending among all cases. There were too few reoffenders to permit comparison among month zero cases, except to confirm the relative weakness of RM2000/v.

Predictor	General reoffending		Violent reoffending	
	All cases	Month zero cases only	All cases	Month zero cases only
Violent reoffending predictors				
RM2000/v	n/a	n/a	0.689 ***	0.694 ***
RM2000/v and OFM	n/a	n/a	0.713 (NC)	n/a
OVP1 static scale	n/a	n/a	0.749 (R)	0.764 (R)
OVP1 static scale and OFM	n/a	n/a	0.760 (NC)	n/a
OGRS4/V	n/a	n/a	0.769 ***	0.779 ***

Note. OFM = polynomial model term for offence-free months. DeLong *et al*. (1988) AUC comparisons were completed for the original predictors (i.e. without OFM addition). (R) = reference predictor. (NC) = comparison not conducted. NS: p>.05. *: p<.05. **: p<.01. ***: p<.001.

Table 8.7 presents actual and predicted probabilities by sentence, and by age and gender. Among juvenile offenders, actual and predicted probabilities were lowest for cautioned offenders. Predicted rates were highest for those receiving custodial offenders, and actual rates were above predicted rates by three and one per cent for general and violent reoffending respectively. Actual reoffending rates were around two per cent below predicted for the small number of juveniles discharged from court. Overall actual and predicted rates were much lower for adult than juvenile offenders. As with juveniles, cautioned offenders had the lowest rates, while imprisoned offenders had the highest predicted rates and positive residuals. Actual rates, especially for general reoffending, were below predicted rates for those serving community and other (i.e. miscellaneous) sentences. Rates were second-highest for the suspended sentence group, with very small residuals.

Among female offenders, actual and predicted probabilities differed by one percent for some groups. Probabilities were highest at the youngest ages, falling rapidly for violent reoffending and more slowly for general reoffending. Indeed, general reoffending probabilities were essentially stable between ages 14 and 34, before falling again. For both types of reoffending, probabilities fell very rapidly from age 50 upwards. Among male offenders, residuals were mostly smaller. Probabilities of general reoffending were lower for 10–11 year olds than other juveniles, while probabilities of violent reoffending differed little across the juvenile age range. After age 16–17, probabilities of both outcomes fell steadily, with very low probabilities among the oldest offenders.

Table 8.7 also presents AUCs, to establish whether offenders can be ranked by relative risk within age, gender and sentence groups. Separation of relative risk will be easier in adult justice than youth justice, with AUCs being several points higher in the former, due to greater variation in criminal history and age. AUCs were higher for all juveniles than for male or female juveniles alone, as gender is an important factor among juveniles due to the difference in male and female age patterns. There is very little variation in risk, and therefore low AUCs, among the very youngest offenders, most of whom are first-time entrants. Intra-age group AUCs, for both male and female offenders, rise with age, with the very oldest age groups mixing many occasional offenders and first-time entrants with small proportions of long-term persistent offenders. The transition from youth to adult justice is not especially affected by the way OGRS4 models age, for two reasons. Firstly, the static nature of age at index date means that aging will not affect an individual's score until they reoffend, so individuals making a transition purely due to their 18th birthday will not be affected. Secondly, as Figure 8.3 shows, while age trends in the late teenage years are strong, the separate age calculations for juveniles and adults only have a modest effect on the pattern between ages 17 and 18.

For juvenile offenders, the AUC measure of relative predictive validity within sentence groups was moderate among cautioned offenders, and good for most other groups. It was very good for general reoffending among those who received custodial sentences. It was also very good for both outcomes among those with 'other' sentences, perhaps because this category covers a diverse range of unusual sentences and therefore has heterogeneous offender characteristics. Most AUCs were several points higher for the corresponding adult offender sentence groups. While most differences between the custodial reference group and the various non-custodial sentence groups were significant among adults, most of these differences were quite small and their statistical significance was more a function of the very large sample sizes.

Table 8.7: Absolute and relative predictive validity: Actual and predicted general and violent reoffending, and comparisons of Area Under Curve (AUC) statistics, by sentence, age and gender

	General reoffending within 2 years of sentence/discharge				Violent reoffending within 2 years of sentence/discharge			
Offender group	**Actual**	**Predicted**	**Residual (Actual – Predicted)**	**AUC**	**Actual**	**Predicted**	**Residual (Actual – Predicted)**	**AUC**
All offenders	**24.4%**	**24.4%**	**-0.0%**	**0.77**	**14.9%**	**15.0%**	**-0.0%**	**0.77**
All juveniles	***22.5%***	***22.5%***	***-0.0%***	***0.72***	***13.1%***	***13.2%***	***-0.1%***	***0.72***
Caution	25.2%	25.3%	-0.1%	0.66***	16.7%	16.5%	0.2%	0.67***
Discharge	49.6%	51.5%	-1.9%	0.72**	35.2%	37.9%	-2.7%	0.70
Fine or compensation	52.9%	52.3%	0.6%	0.73**	35.1%	37.2%	-2.0%	0.71
Community	48.1%	47.4%	0.7%	0.73***	34.1%	34.0%	0.1%	0.71
Suspended	See note				See note			
Custodial	63.8%	60.6%	3.2%	0.78 (RC)	46.6%	45.8%	0.9%	0.73 (RC)
Other	57.4%	55.7%	1.7%	0.81	43.1%	41.6%	1.5%	0.78*
All adults	***31.9%***	***31.9%***	***0.0%***	***0.78***	***21.8%***	***21.7%***	***0.1%***	***0.77***
Caution	15.3%	15.1%	0.1%	0.73***	9.3%	9.3%	-0.1%	0.73***
Discharge	27.0%	27.5%	-0.5%	0.76***	16.1%	16.1%	-0.1%	0.76**
Fine or compensation	21.6%	21.9%	-0.3%	0.78	12.3%	12.5%	-0.2%	0.80***
Community	27.3%	28.0%	-0.7%	0.75***	15.8%	16.2%	-0.3%	0.75***
Suspended	33.9%	33.8%	0.1%	0.76***	19.8%	19.9%	-0.1%	0.75***
Custodial	39.7%	37.5%	2.2%	0.78 (RC)	22.5%	21.4%	1.1%	0.77 (RC)
Other	26.6%	28.3%	-1.8%	0.77	15.4%	16.6%	-1.2%	0.77

Offender group	General reoffending within 2 years of sentence/discharge				Violent reoffending within 2 years of sentence/discharge			
	Actual	Predicted	Residual (Actual – Predicted)	AUC	Actual	Predicted	Residual (Actual – Predicted)	AUC
Female offenders	***16.1%***	***16.1%***	***-0.1%***	***0.76***	***9.2%***	***9.2%***	***0.0%***	***0.77***
All juveniles	18.9%	18.6%	0.3%	0.69	12.9%	12.7%	0.2%	0.72
All adults	15.0%	15.1%	-0.2%	0.78	7.8%	7.9%	-0.1%	0.78
10–11	22.6%	24.0%	-1.4%	0.53***	15.9%	16.5%	-0.5%	0.57***
12–13	21.5%	20.3%	1.2%	0.65***	14.7%	13.9%	0.7%	0.70***
14–15	17.6%	18.0%	-0.5%	0.68***	12.3%	12.6%	-0.3%	0.72***
16–17	18.9%	18.1%	0.8%	0.70***	12.4%	12.0%	0.5%	0.73**
18–20	16.8%	16.8%	0.0%	0.73**	10.5%	10.2%	0.4%	0.75
21–24	15.9%	16.9%	-1.0%	0.76 (RC)	8.2%	9.1%	-0.8%	0.76 (RC)
25–29	17.8%	18.0%	-0.2%	0.79***	8.5%	8.8%	-0.3%	0.78
30–34	17.0%	17.0%	-0.1%	0.79***	8.2%	8.2%	-0.1%	0.76
35–39	15.7%	15.3%	0.4%	0.77*	8.0%	7.7%	0.3%	0.77
40–49	11.9%	12.0%	-0.1%	0.78**	6.1%	6.0%	0.1%	0.78*
50–59	6.9%	7.2%	-0.3%	0.77	3.1%	3.2%	-0.1%	0.82**
60+	3.4%	3.7%	-0.3%	0.82	1.4%	1.3%	0.1%	0.86**
Male offenders	***26.8%***	***26.8%***	***-0.0%***	***0.76***	***16.6%***	***16.6%***	***-0.1%***	***0.76***
All juveniles	37.7%	37.8%	-0.1%	0.69	25.8%	25.7%	0.0%	0.69
All adults	24.3%	24.3%	0.0%	0.77	14.5%	14.5%	-0.1%	0.77
10–11	33.2%	32.2%	1.0%	0.58***	25.2%	24.2%	1.0%	0.58***
12–13	37.7%	37.0%	0.6%	0.65***	26.7%	26.0%	0.7%	0.64***
14–15	38.1%	38.3%	-0.2%	0.69***	25.8%	25.9%	-0.1%	0.68***
16–17	38.0%	38.3%	-0.3%	0.72***	25.4%	25.7%	-0.2%	0.71***
18–20	32.9%	32.5%	0.4%	0.73***	21.3%	21.1%	0.2%	0.73***
21–24	28.4%	29.2%	-0.8%	0.75 (RC)	17.4%	17.9%	-0.5%	0.75 (RC)
25–29	26.6%	26.5%	0.1%	0.76***	15.2%	15.3%	-0.1%	0.75
30–34	24.9%	24.1%	0.8%	0.76*	14.1%	13.5%	0.6%	0.74
35–39	22.4%	22.1%	0.3%	0.76*	12.7%	12.7%	0.1%	0.75
40–49	18.0%	18.4%	-0.4%	0.76	10.4%	10.7%	-0.3%	0.76**
50–59	10.7%	11.2%	-0.5%	0.76	5.7%	6.2%	-0.4%	0.79***
60+	5.8%	5.4%	0.4%	0.77	3.2%	2.8%	0.4%	0.81***

Note. RC = reference category for T-test AUC comparisons. T-test comparison with reference category AUC is significant at: * $p<.05$; ** $p<.01$; *** $p<.001$. All groups had $n > 1000$, except that 48 juveniles were recorded with suspended sentences. Statistics were therefore not produced for this group.

Table 8.8 gives decile-wise information on the distribution of OGRS4/G and OGRS4/V scores, as used in the Hosmer-Lemeshow test. For the general predictor, reoffending rates were fractionally above predicted for each of the six lowest deciles, one percent below predicted for the 8th and 9th deciles, and one percent above predicted for the 10th decile. These suggest a modicum of difficulty in calibrating risk levels for those well over the average risk level. However, the actual reoffending rates associated with each decile were still very well separated from those of adjacent deciles. For violent reoffending, no residuals were higher than 0.5%, and the top decile produced more reoffenders than the entire bottom three-fifths of the distribution (i.e. deciles 1 to 6 combined); this was almost true for general reoffending.

Table 8.8: Actual and predicted probability of general and violent reoffending by predictive score decile

Decile	General reoffending within 2 years of sentence/discharge			Violent reoffending within 2 years of sentence/discharge		
	Actual	Predicted	Residual (Actual – Predicted)	Actual	Predicted	Residual (Actual – Predicted)
1	4.3%	4.2%	0.1%	1.7%	2.0%	-0.3%
2	7.1%	7.0%	0.2%	3.4%	3.6%	-0.2%
3	9.7%	9.6%	0.1%	5.1%	5.2%	-0.1%
4	12.9%	12.7%	0.1%	7.0%	6.9%	0.1%
5	17.3%	16.8%	0.5%	9.5%	9.2%	0.3%
6	21.9%	21.7%	0.2%	12.7%	12.2%	0.5%
7	27.4%	27.5%	-0.1%	16.2%	15.8%	0.3%
8	33.6%	35.0%	-1.3%	20.5%	20.6%	-0.1%
9	44.2%	45.7%	-1.4%	27.7%	28.2%	-0.5%
10	65.7%	64.2%	1.5%	45.5%	45.9%	-0.4%

Note. N = 60,396 for decile 1, 60,397 for deciles 2–9, 60,389 for decile 10.

8.4 Implications

The application of offence-free time and other innovations in the coding of age and criminal history lead to improvements in the prediction of both general and violent reoffending, compared with OGRS3. The offence-free time concept is only useful when considering caseloads sanctioned/discharged at different points in time, such as the NOMS caseload, and therefore the benefits of OGRS4 are greatest when used with such mixed caseloads. However, compared with OGRS3, worthwhile improvements in the prediction of violent reoffending, though only small improvements for general reoffending, are also found among just-sanctioned/discharged offenders (those with zero offence-free months).

The coding of the predictors is more complex than that of OGRS3. However, this does not impact operational users who need only enter the same few primary risk factors required in OGRS3. Rather, it illustrates the degree of fine-tuning which is required to achieve incremental improvements in the prediction of proven reoffending.[134]

Compared with previous studies (Howard *et al*., 2009; Howard, 2009), OGRS3's AUCs for general reoffending are relatively lower, whereas those for violent reoffending are relatively high. This indicates that violence risk was more dispersed in this sample than previous samples, perhaps reflecting a high proportion of never-violent offenders among those receiving cautions and other non-NOMS disposals, compared to a small group with extensive violence history and therefore high predicted probabilities of violent reoffending. By contrast, the homogeneity of the sample in terms of overall criminal history may have given reduced scope to distinguish reoffenders and non-reoffenders for any offence. These shifts in risk distribution are a consequence of the major change in sampling strategy between OGRS3 and OGRS4, where the former optimised prediction for those starting sentence/discharge while the latter optimises prediction for something closer to the whole caseload at a given point in time.

The predictors have scope to be used in settings where OGRS3 is not currently used, among offenders with non-conviction sanctions and in youth justice. Expanding the use of OGRS4 would require the development of user guidance and possibly training. Users who are already familiar with OGRS3 could be issued with more limited guidance covering the improved validity, revisions to offence categories, the offence-free time element, and any subsequent revisions to risk categories. The offence-free time element may be the most challenging, but importantly has the potential for improving the targeting of supervision and treatment resources (Howard, 2011).

In practice, risk predictors are often used in categorised form, i.e. with predictions reported as low, medium, high or very high. These categories will need to be calibrated carefully, recognising that due to the introduction of the offence-free time element, mean predicted scores are lower than those for OGRS3. Further work will also need to be undertaken to create the next-one-year predicted probabilities which correspond to the two-year probabilities. As set out earlier in this chapter, regular recalibration of actuarial predictors is important to ensure that they reflect changing levels and patterns of reoffending. Further recalibration of OGRS4 is summarised in Chapter 13, based upon a dataset of offenders on the NOMS community caseload in 2010. This recalibration does not change the key findings.

[134] Given the difficulties achieving improvements by the use of still more complex methods such as neural networks (Yang, Liu and Coid, 2010), any further improvements may only come from additional innovations on secondary risk factor coding within logistic regression models. Such models are used to measure performance in reducing reoffending (e.g. Ministry of Justice, 2012) and do include more complex terms, but these require more primary risk factors to be collected, and therefore are only suitable for use in central statistical activities rather than operational practice.

9. OGP2 and OVP2: the revised OASys predictors

This chapter reports on the development of version 2 of OGP and OVP, setting out the following key points:

- Following feedback from OASys users, the second iteration of OGP predicts all proven reoffending.
- OGP2 and OVP2 have the same static risk factors as those used in OGRS4/G and OGRS4/V (see Chapter 8), although these factors are scored differently.
- As with OGRS 4, the new models include an 'offence-free time' element.
- Dynamic risk factors in both predictors include accommodation, employability, intimate partner relationships, the type of drug used, alcohol misuse, impulsivity, temper control and problem solving skills. OGP2 also includes frequency of drug misuse and pro-criminal activities and attitudes.
- Improvements in the prediction of both general and violent reoffending resulted from the application of offence-free time and other innovations in the selection and coding of risk factors. Version 2 better distinguishes reoffenders from non-reoffenders, and better calibrates actual and predicted reoffending rates for certain offender groups: the highest- and lowest-risk, the oldest and youngest, and females (see Chapter 3).
- The predictors could be introduced in a revision to OASys, accompanied by user guidance.

9.1 Context

Within OASys, likelihood of recidivism is estimated using actuarial (mathematical) scoring rules rather than clinical judgement, although RoSH is rated clinically. OASys includes both static (e.g. age, gender, criminal history) and dynamic risk factors (e.g. accommodation, substance misuse, thinking skills, attitudes), and both are used in the actuarial risk predictors. Scores on the OASys Violence Predictor v.1 (OVP1; Howard, 2009) estimate with reasonable predictive validity the offender's likelihood of proven recidivism involving a broad group of violence-related offences, and also identify those most likely to commit homicide and wounding with intent, the most serious non-sexual violent offences. Scores on the OASys General reoffending Predictor v.1 (OGP1; Howard, 2009) estimate with good predictive validity the likelihood of proven recidivism for most non-sexual, non-violent offences. The inclusion of dynamic factors means that offenders' scores can change over time, if and when they are reassessed using OASys.

OGP1 and OVP1 were developed by Howard (2009) and implemented in August 2009. For OGP1, the OGRS3 score (Howard *et al*., 2009) was used as the static risk component of the score, whereas OVP1 includes a bespoke selection and weighting of demographic, general and violent offending history factors for its static risk component. Both predictors produce a 100-point score, with 60 points for static and 40 points for dynamic risk factors. Subsequent to the Howard (2009) study, alternate versions of the predictors were produced to make it possible to score them using the more limited sets of items in standard and full OASys. Table 9.1 summarises these risk weightings and items. Appendix B gives more detail of the individual items used.

Table 9.1: Static and dynamic risk factors scored in OGP1 and OVP1

Risk factor[135]	**OASys items used in standard**	**Additional OASys items used in full**	**Weight in OGP1**	**Weight in OVP1**
OGRS3 score	n/a	n/a	60	--
Sanctions for violent offences			--	25
Sanctions for non-violent offences			--	5
Not first sanction ever			--	5
Age at current community order/custody discharge			--	20
Male			--	5
Total weight of static risk factors			***60***	***60***
Does not recognise impact of offending	2.6	None	0	4
Accommodation	3.3	3.4, 3.5, 3.6	5	4
Employability	4.2	4.3, 4.4, 4.5	5	6
Regular activities encourage offending	7.2	None	5	0
Drug misuse	8.4, 8.8	8.5, 8.6, 8.9	15	0
Alcohol misuse	9.1, 9.2	None	0	10
Current psychiatric treatment, or treatment pending	10.7	None	0	4
Thinking and behaviour	11.6, 11.7, 11.9	11.5	5	0
Temper control	11.4	None	0	6
Attitudes	12.1	12.4, 12.5, 12.8	5	6
Total weight of dynamic risk factors			***40***	***40***
Total weight of all risk factors			***100***	***100***

For each predictor, the 100-point score is then transformed into predictors of proven reoffending committed within one and two years of the start of community sentence or discharge from custody. Subsequent research (see Chapter 4) has shown that the predictors include genuinely dynamic risk factors. Their predictive validity is thus greater when used in practice as offenders are reassessed over time.

OGP1 and OVP1 were produced using offender assessments completed between 2002 and 2004, and a recalibration using more recent assessments could account for changes in offender risk/need patterns and changes in levels of recidivism (Ministry of Justice, 2013a). In updating the predictors, account can be taken of recent research on reoffending patterns and the properties of OGP and OVP (see Chapters 3 to 5). Further improvements may be possible as new insights on the possible

135 Howard (2009) erroneously stated the age item as “age at current conviction”.

structure of static and dynamic risk can be tested. The aims of this paper were therefore to investigate possible improvements to OGP and OVP in the following five areas:

1. ***Identifying reductions in the hazard of reoffending over time.*** A recent dataset of offenders was explored to confirm that accounting for offence-free time improved prediction. Chapter 8 introduced offence-free time to OGRS4, and a similar update to OGP and OVP would keep the predictors aligned. Introducing offence-free time, if it improves prediction, will optimise the value of OGP and OVP in real-world situations where reassessments do take place, and where offenders' resourcing levels can be re-evaluated if evidence suggests that they have become less likely to reoffend.
2. ***Optimising the use of dynamic risk factors to predict reoffending.*** Recent data was analysed to determine which dynamic risk factors were most helpful in predicting all and non-sexual[136] violent proven reoffending, controlling for static and offence-free time risk factors. Chapter 4 demonstrated that the social/personal items included in OGP and OVP differ in the extent to which they are genuinely dynamic risk factors. The scores of genuinely dynamic risk factors are prone to change as OASys is reassessed over the course of probation supervision, and these score changes are incrementally predictive of reoffending. As with offence-free time, the inclusion of genuinely dynamic risk factors increases the value of OGP and OVP because it allows offender managers to adjust resources consistently in response to falls – and, less often, rises – in likelihood of reoffending.
3. ***Constructing statistical models to predict all reoffending, and non-sexual violent reoffending.*** These models were developed to predict 'next two years' reoffending: reoffending either in the two years following community sentence or discharge from custody, or in the two years following a subsequent offence-free period ranging from one month to three years. This aims to correctly prioritise offenders with different amounts of offence-free time, unlike the wholly static, 'first two years', OGP1 and OVP1 scores. The use of all rather than non-violent reoffending responds to stakeholder feedback indicating that score interpretation and usage would be simpler if OGP predicted all rather than non-violent recidivism. This would bring OGP into line with OGRS: both the old (version 3) and new (version 4) OGRS instruments feature a predictor of all reoffending. As OGRS4 introduced a violence predictor, the opportunity exists to align the two families of predictors, with general (i.e. all) and violence predictors in both.
4. ***Updating the user-friendly OGP and OVP risk predictors.*** It is important to update actuarial predictors every few years, as the correlates and overall prevalence of reoffending may change, and the new information on offence-free time and the use of dynamic risk factors presents an opportunity to make such an update. As such, the models constructed above were used to create OGP2 and OVP2. As with OGP1 and

[136] NOMS uses the static risk predictor Risk Matrix 2000/S (Thornton, 2007) to assess proven sexual reoffending risk among offenders with a known history of sexual offending.

OVP1, the risk factors which constitute the new predictors were weighted to a total of 100 points, to allow assessors to understand the balance of offenders' risk factors. To optimise efficiency in offender assessment practice, the static risk factors used were identical to those used in OGRS4. Variants of the models then tested whether OGP2 and OVP2 should incorporate the OGRS4 scores themselves or enter each OGRS4's risk factors as separate model terms.

5. ***Validating the new OGP2 and OVP2 risk predictors.*** The new predictors were validated on a further sample of offenders, comparing them with OGP1 and OVP1, and checking their validity for offenders of different age, gender and ethnicity.[137] The models were also validated against OGRS.

9.2 Approach

Sample

A list of offenders assessed using OASys by 31 March 2008 was created, filtering out assessments with missing dynamic risk factor data or RoSH ratings, and ensuring that each offender was only represented once during each period of contact with the criminal justice system. This list was submitted to the Ministry of Justice's (MoJ's) Police National Computer (PNC) research database in December 2010. After filtering out: (i) those whose index conviction (i.e. the conviction for which OASys was being completed) could not be identified on the PNC; (ii) those whose assessment was not within three months of their community sentence or discharge from custody; and (iii) those for whom OGRS3, OGP or OVP scores could not be calculated,[138] 156,837 cases had the potential to be included in the two-year follow-up.

This chapter focuses upon '***next-two-year***' follow-ups rather than traditional '***first-two-year***' follow-ups (see Section 8.2). This approach comes closer to maximising predictive validity across NOMS's community caseload, rather than its commencements. As the maximum length of Community Orders is three years, cases were included in the next-two-year follow-up sample when their follow-up start date was no more than three years after their date of sentence/discharge. Follow-ups commencing in March 2008 allowed at least seven months (due to December 2010 PNC submission date) for offences committed during the follow-up to result in a PNC-recorded conviction. Offenders who were sentenced/discharged between March 2005 and March 2008 were traced until the equivalent date in March 2008. Among those with multiple sentences/discharges in this period, only the latest sentence/discharge was retained.[139] These offenders were retained in the sample if they did not

---

[137] The validity of OVP2 was also checked for the outcome of proven homicide/wounding reoffending, ensuring that it improves prediction of the most harmful non-sexual violent recidivism.

[138] Due to missing date of birth or apparent convictions aged under 10, or missing data on OGP1 or OVP1 items.

[139] When reoffending occurs, predictors of reoffending should be rescored, and offenders will often start a new period of contact with NOMS which can lead to new start of sentence/discharge OASys assessments. Therefore, despite OVP predicting violent reoffending only, proven reoffending of any type censored the pre-March 2008 follow-up period, avoiding duplication in the sample and ensuring correct scoring of static factors. By contrast, the next-two-year predictions relate to whether an offence of the relevant type is committed at any point in the next two years using the information available at the time the prediction is made. A non-violent offence partway through the next-two-year follow-up does not therefore invalidate the OVP prediction, and next-two-year follow-ups are not censored on this basis.

commit a proven reoffence of any type, nor were imprisoned for any offence,[140] prior to the March 2008 equivalent date. Potential March 2008 follow-up start dates and offence-free time values were then calculated for each case. Dates of first reoffending, reimprisonment and recall to custody were checked, starting from the 2005–08 sentence/discharge date.

All cases were included in the exploratory survival analysis. Cases were split into multiple observations covering the periods between successive assessments, with their dynamic risk factors varying between assessments. These were treated as time-dependent covariates (discussed more thoroughly in Chapter 4) in statistical modelling. The 152,141 cases in the survival period (i.e. prior to their potential follow-up start date) had 337,973 assessments in this period, including their initial assessments.

A further 4,696 offenders commenced community sentences or were discharged from custody in March 2008, and were not included in exploratory survival analysis. For these offenders, their initial assessment for the next-two-year follow-up was naturally the assessment already linked to their sentence/discharge. For all other offenders, their final assessment in the survival period was copied across to become their initial assessment for the next-two-year follow-up. Next-two-year general and violent reoffending statuses were calculated for all those in the next-two-year sample. Offenders were removed from this sample if they were imprisoned or recalled at a date which was both within the two-year period and prior to any proven reoffending.

Many offenders were reassessed with OASys during the next-two-year period, and reassessments prior to proven reoffending were taken into account in estimating the model parameters. For statistical and computational reasons, these cases' changing scores were not treated as time-dependent covariates, unlike the scores of similar cases in the survival period and the studies presented in Chapters 4 (change analysis) and 10 (sexual reoffending predictor).[141] Instead, one summary observation was created per case, in which dynamic risk scores were weighted according to the number of days of the 730-day-long follow-up they represented, with the remainder of a reoffender's follow-up always allocated to their final assessment.[142] This allowed scores to be updated as they

[140] It is possible for an offender to be imprisoned without committing a new offence, as the result of a pseudoreconviction. Pseudoreconvictions are offences committed before the start of the period of interest but brought to justice after. Information on periods on remand in custody, which could also disrupt a reoffending follow-up, was not available.

[141] In Howard (2011), Cox proportional hazards regression was used to estimate the effect of time-dependent covariates on reoffending. Dynamic risk factors were separated into an effectively static portion – the score at initial assessment – and a wholly time-dependent portion – the change from initial to current score. It is not possible to use Cox regression with the next-two-year follow-up, as it does not model the baseline hazard and therefore does not produce predicted probabilities. Accelerated failure time regression using a Weibull process was used in Chapter 10 in order to overcome this problem. Dummy cases were used to predict two-year reoffending, despite the variable follow-up period in this sample. However, Weibull models are computationally intensive to produce, making this procedure unviable using the very large samples and large numbers of feasible predictive risk factors available in this study. (The Chapter 10 study purposefully restricted the number of risk factors considered to limit statistical power concerns related to the small number of reoffenders (Steyerberg, 2009).)

[142] For example, Offender A scored 2 on a risk factor at their initial assessment, and 0 at a reassessment conducted after 73 days. Offender A did not reoffend. Their weighted score on this risk factor was 0.2 (= ((2*73)+(0*657))/730). Offender B scored 1 at initial assessment and 2 at a reassessment after 146 days, and reoffended between the 146- and 730-day point. Their weighted score on this risk factor was 1.8 (= ((1*146)+(2*584))/730).

would be as reassessments occurred over the course of real two-year follow-ups,[143] while using a weighting system to ensure that each offender was effectively represented only once when constructing the predictors.

The final next-two-year sample included 109,319 cases (i.e. those 4,429 of the 4,696 sentenced/discharged in March 2008, and those 104,890 of the 106,459 March 2005 to February 2008 cases, who neither reoffended, were recalled nor were imprisoned for a pseudoreconviction prior to their potential March 2008 start date). The sentence breakdown was as follows (sentence details were unrecorded for 5%):

- 23% on licence from a custodial sentence;
- 50% on Criminal Justice Act 2003 (CJA 2003) Community Orders;
- 17% on Suspended Sentence Orders; and
- 9% on pre-CJA 2003 Community Sentences.

The sample included few offenders with custodial sentences of under 12 months, nor with non-rehabilitative community sentences, as OASys is not routinely used with these offenders. Demographic details of the final next-two-year sample are included in the results section. The 109,319 cases had 118,546 further assessments during the next-two-year follow-up, preceding their date of first proven reoffending.

The final next-two-year sample was randomly split into construction (two-thirds of cases) and validation (one-third of cases). The construction sample contained a total of 152,116 assessments for 73,098 offenders; the validation sample contained a total of 75,749 assessments for 36,221 offenders.

Procedure

Previous sanctions, proven reoffending and recalls

Records of offenders' known criminal history and proven reoffending were extracted from the MoJ's PNC research database. Recall dates were identified by matching offenders with an extract from NOMS's Public Protection Unit Database. Reoffending, imprisonment and recall prior to the next-two-year follow-up start date was checked as described above, to create the next-two-year sample. In the next-two-year follow-up, proven reoffending comprised offences committed between each offender's follow-up start date and the same date two years later.

[143] There is arguably a problem with this approach for the OVP follow-up, as the reassessment process was treated as ceasing when any proven reoffence occurred. If this was a non-violent offence, then a confusing situation was encountered where the OVP prediction was still 'live' for the purposes of this project and indeed from the practice perspective of a next-two-year prediction made on a prior date, yet in practice the offender would have been rescored (on static and offence-free as well as dynamic risk factors). This situation would have been easier to represent in a Weibull model, but this was not practical in this study. In the absence of such a solution, the cessation of the reassessment process upon any reoffending was adopted as a pragmatic and relatively simple solution; the use of any reassessments in predictive model development is still a considerable step forward, and the overall approach represents the dynamic nature of risk factors much better than that taken in the development of OVP1.

Proven ***violent reoffending*** is classified in OVP1 as any proven reoffending involving offence(s) of homicide and assault, threats and harassment, violent acquisitive offences (robbery and aggravated burglary), public order, criminal damage and/or weapon possession. Contact sexual offences were excluded from the classification as they were shown to be unlikely to aid such prediction. It is reiterated that this classification is designed to maximise predictive validity and does not imply that sexual offences are not harmful. OVP1 uses this classification to determine which offences count as previous violent and non-violent sanctions, and which count as proven violent reoffending. OVP2 will do the same, with the amendment that arson offences can be included (see Chapter 5). OGRS4/V uses an identical classification to OVP2.

Selection of static and dynamic risk factors

Static risk factors were coded from PNC data. The choice of static risk factors was constrained to ensure that all could be coded swiftly by practitioners or administrative staff on the basis of summary printouts of individual offenders' demographics and criminal histories. Such printouts are routinely available from the operational PNC. Dynamic risk factors were selected from the offending-related factors within OASys. In this step, and all subsequent steps, the full rather than standard versions of OGP1 and OVP1 were used. This ensures that later comparisons between version 1 and 2 were conservative, with the 'best' versions of the existing predictors always being used.

Cox regression modelling with time-dependent covariates was conducted for exploratory analysis of the survival sample. This was used to test different coding options for static and dynamic risk factors, including:

- the selection of the optimal Copas rate[144] for general reoffending;
- whether a similar rate could usefully be created either for custodial sanctions or sanctions for violent offences; and
- the creation of indicator variables for second-, third-, fourth- and fifth-time entrants.

This exploratory analysis was conducted upon the survival dataset as it was a very large sample which covered a different time period to the next-two-year sample. Using this sample to test coding possibilities, and then double-checking the value of new items on the next-two-year construction sample, minimised the possibility of overfitting the models by introducing unproductive innovations.

[144] The first offending rate of this type was designed by Professor John Copas for OGRS1 (Copas and Marshall, 1998).

When provisional logistic regression models had been selected, using the construction sample, user-friendly versions were created. While logistic regression models are suitable for use by researchers, and could be scored for practitioners in the OASys IT application, assessments with dynamic elements should allow practitioners to consider what steps can be taken to reduce the likelihood of reoffending. When the offender's score is expressed in points and visible to the practitioner, and clearly divided between the various risk factors, areas of strength and weakness can be identified, and their impact can be communicated to the offender. This is more important with offence-free time and dynamic risk factor scores, as these scores drive changes in offenders' predicted likelihoods of reoffending over the course of a single sentence, unlike static scores. This is why OGP1 and OVP1 are expressed as 100-point scores as well as percentage likelihoods, and this method was repeated for OGP2 and OVP2.

The method for obtaining scores is set out in detail in Appendix I, though one important point must be mentioned here. The weights given to the three types of risk factor (i.e. static, dynamic, offence-free time) were obtained by comparing their ranges – the differences between maximum and minimum possible regression coefficient sums. However, a very small proportion of offenders have extremely high scores on static risk factors. To avoid these outliers distorting model fit, extreme scores were truncated before the summary scoring of the three risk factor types.

In order to keep scoring uncomplicated, items with one-point weights were removed and the models refitted and rescored. Where possible, single items were used, rather than multi-item scales. This should allow OGP2 and OVP2 to be used as widely as possible in the future, recognising that resource pressures could prompt revisions to OASys content and/or use of shorter versions of OASys. The weighting of dynamic items was made simple for practitioners to apply and understand, wherever possible. The use of items not currently included in 'standard' OASys (which does not have all the items required for the criminogenic need scores – see Chapter 11) was permitted, but where intercorrelations made items essentially interchangeable,[145] items included in standard OASys were preferred.[146]

The process described above was run twice for each outcome (i.e. for the G and V predictors). In one version, labelled 'no-OGRS', static risk factors were fitted as a direct part of the predictor. In the other version, labelled 'with-OGRS', OGRS4/G or OGRS4/V scores were used to provide the static and O-FT elements of the predictors. Utilising OGRS scores offers several advantages over static/O-FT/dynamic models which are entirely separate from the OGRS family of predictors.[147]

[145] There are strong correlations between many of the items comprising OASys's criminogenic need scales (see Chapter 11).
[146] Interactions between static risk factors, offence-free time and/or dynamic risk factors were also trialled, entering the three points scores into further logistic regression models. These did improve model fit slightly, suggesting that dynamic risk factors have more predictive value for offenders with low levels of static risk, but were judged too complex to implement, so were not taken forward.
[147] Frontline staff would face a less complex task in understanding the vagaries of predictions themselves and in explaining them to offenders. The predictors will all need to be introduced into operational IT systems, and systems testing would be simplified by using a smaller number of algorithms. Future research would also be simpler, and detailed discussion of the four predictors both inside and outside NOMS would be easier if fewer separate models existed.

However, nesting logistic regression models within one another may bias prediction (Steyerberg, 2009), and therefore model validity tests were deemed vital in establishing whether 'with-OGRS' predictors could be used.

Checks of absolute predictive validity (actual versus predicted recidivism rates) and relative predictive validity (using Area Under Curve (AUC) statistics) were conducted for the user-friendly versions of the competing predictors. One set of absolute comparisons checked whether actual rates equalled predicted rates across the range of predicted rates, while another set of absolute and relative comparisons checked whether this was true for offenders of different age, gender, ethnicity and sentence type. After these comparisons, a decision was made on whether the former ('no-OGRS') or latter ('with-OGRS') version of the predictors should be taken forward. The validation sample was used for the age and gender comparisons (as age and gender are among the risk factors used in the predictors) but the entire next-two-year sample was used for the ethnicity and sentence comparisons (as the numbers in some groups were small, and neither ethnicity nor sentence were considered for inclusion as risk factors).

As the predictive models for all and violent reoffending were generated independently from one another, it was possible that some offenders could receive a predicted probability of violent reoffending which was higher than their predicted probability of any reoffending. Given that each offender's true probability of violent reoffending must be no higher than their true probability of any reoffending, it would be unsatisfactory for them to receive a predicted probability of violent reoffending which exceeds their predicted probability of any reoffending. Therefore, it was also checked whether in some cases the OGP2 prediction needed to be increased to equal the OVP2 prediction.

Model validation

To test relative risk, AUC statistics were calculated for validation sample cases. As the predictive scores were correlated, all significance tests for differences in AUCs were conducted using a non-parametric comparison developed by DeLong *et al*. (1988).[148] The AUC of OGP2 was compared with those of OGRS3, OGRS4/G, and OGP1, as predictors of all reoffending. For violent reoffending, the AUCs of OVP2, OGRS3, OGRS4/V, OVP1 and the established static violence risk predictor Risk Matrix 2000/v (RM2000/v; Thornton *et al.*, 2003; Thornton, 2007) were compared.[149]

In practice, much of the value of OGP and OVP derives from their ability to sort offenders into low, medium, high and very high categories, which are reported on the OASys summary sheet and (for OVP) its RoSH component. The bandings also affect resource allocation and offending behaviour programme targeting. Therefore, OGP2 and OVP2 were also compared with the other predictors by grouping the OGP1 and OVP1 predictors according to their existing categories and then creating

[148] This test is available in SAS software version 9.2, which was used in all data preparation and analysis.
[149] In order to maximise its predictive validity, RM2000/v was treated in its 0-8 raw form rather than its four-group categorised form.

identically-sized groups when ranking offenders by the other predictors. When contrasting actual and predicted reoffending, three pairs of sensitivity and specificity statistics were calculated for each predictor, as predictor categories can be used as the basis of three decision thresholds (i.e. prioritising non-low offenders, prioritising high/very high offenders, and prioritising very high offenders only).[150]

9.3 Results

Identifying reductions in the hazard of reoffending over time

A simple exercise was undertaken to establish the value of the offence-free time (O-FT) item, using recidivism data linked to each offenders' initial assessment (*N*=156,837). The data were transformed to create duration logistic regression models, which include one observation for each quarter in which an offender could be followed up.[151] Figure 9.1 compares two duration logistic models fitted for any proven reoffending, one using only OGP1 scores as a predictor, and the other using OGP1, the quarter count and the squared quarter count as predictors. Figure 9.2 compares a similar pair of models fitted for proven violent reoffending, using OVP1. Table 9.2 presents the statistical models used.

Table 9.2: Duration logistic regression models of reoffending, using OGP1/OVP1 score and offence-free time

	Logistic regression coefficient for all reoffending (standard error)		Logistic regression coefficient for violent reoffending (standard error)	
Risk factor	**Static/dynamic**	**S/d plus offence-free time**	**Static/dynamic**	**S/d plus offence-free time**
OGP1 100-point score	0.0448 (0.00207)	0.0415 (0.00209)	n/a	n/a
OVP1 100-point score	n/a	n/a	0.0645 (0.000423)	0.0611 (0.000426)
Quarter	n/a	-0.1333 (0.00310)	n/a	-0.0979 (0.00441)
Quarter, squared	n/a	0.00348 (0.00203)	n/a	0.00191 (0.000291)
Constant	*-4.6612 (0.0107)*	*-3.93010 (0.0143)*	*-6.2394 (0.0203)*	*-5.6273 (0.0246)*

Note. All terms statistically significant at *p*<.001. S/d = static/dynamic. N = 152,792 offenders (quarter 1), 1,328,472 offender-quarters (quarters 1 to 20 combined).

The residuals – the differences between actual and predicted reoffending in each quarter – were examined. If offence-free time is a worthwhile risk factor, then the models containing only OGP1 or

[150] Sensitivity equals the proportion of reoffenders who are predicted to reoffend (i.e. whose scores are above the decision threshold). Specificity equals the proportion of non-reoffenders who are predicted to not reoffend (i.e. whose scores are below the decision threshold).

[151] For example, an offender who reoffended in the 9th quarter of follow-up would have nine observations, with quarter counts numbered 1 to 9 and reoffending status coded 'No' for quarters 1 to 8 and 'Yes' for quarter 9. An offender who did not reoffend in quarters 1 to 8 and had a censoring event (recall, reimprisonment for a pseudoreconviction or reaching 3 June 2010) in the 9th quarter would have eight observations, with quarter counts numbered 1 to 8 and all reoffending statuses coded 'No'. Due to censoring events, a small proportion of the 156,837 offenders could not be followed up even into the first quarter. Due to the 3 June 2010 condition, the maximum possible follow-up length was 20 quarters.

OVP1 scores would produce clinically and statistically significant residuals while those also involving the quarter count would have very small residuals. This was indeed the case, although the patterns of predicted hazards require some explanation. In models only using the OGP1 or OVP1 score, without additional model terms to account for offence-free time, predicted hazards fell somewhat over time. This occurred because offenders with higher scores tended to swiftly reoffend, be recalled or be imprisoned for pseudoreconvictions, with those staying in the sample in later quarters having lower mean scores. However, these modest trends in OGP1 and OVP1 scores captured less than half of the true changes in reoffending hazards over offence-free time. Using the models lacking O-FT, the incidences of both types of reoffending were greatly underestimated in quarters 1 and 2 (i.e. months 1–3 and 4–6), moderately underestimated in quarter 3, moderately overestimated in quarter 6, and greatly overestimated from quarter 7 onwards. When simple and quadratic O-FT terms were incorporated into the predictive models, these model residuals shrank so much that they ceased to be clinically significant.[152] These results confirm that including offence-free time should improve risk prediction.

[152] The mean absolute differences between predicted and actual quarterly reoffending rates were 1.6% (all reoffending, no O-FT term), 0.3% (all reoffending, with O-FT), 0.7% (violent reoffending, no O-FT term) and 0.1% (violent reoffending, with O-FT).

Figure 9.1: Prediction of any proven reoffending by OGP1 score and offence-free time

(Hazard = chance of reoffending in this 3-month period IF no reoffending previously)

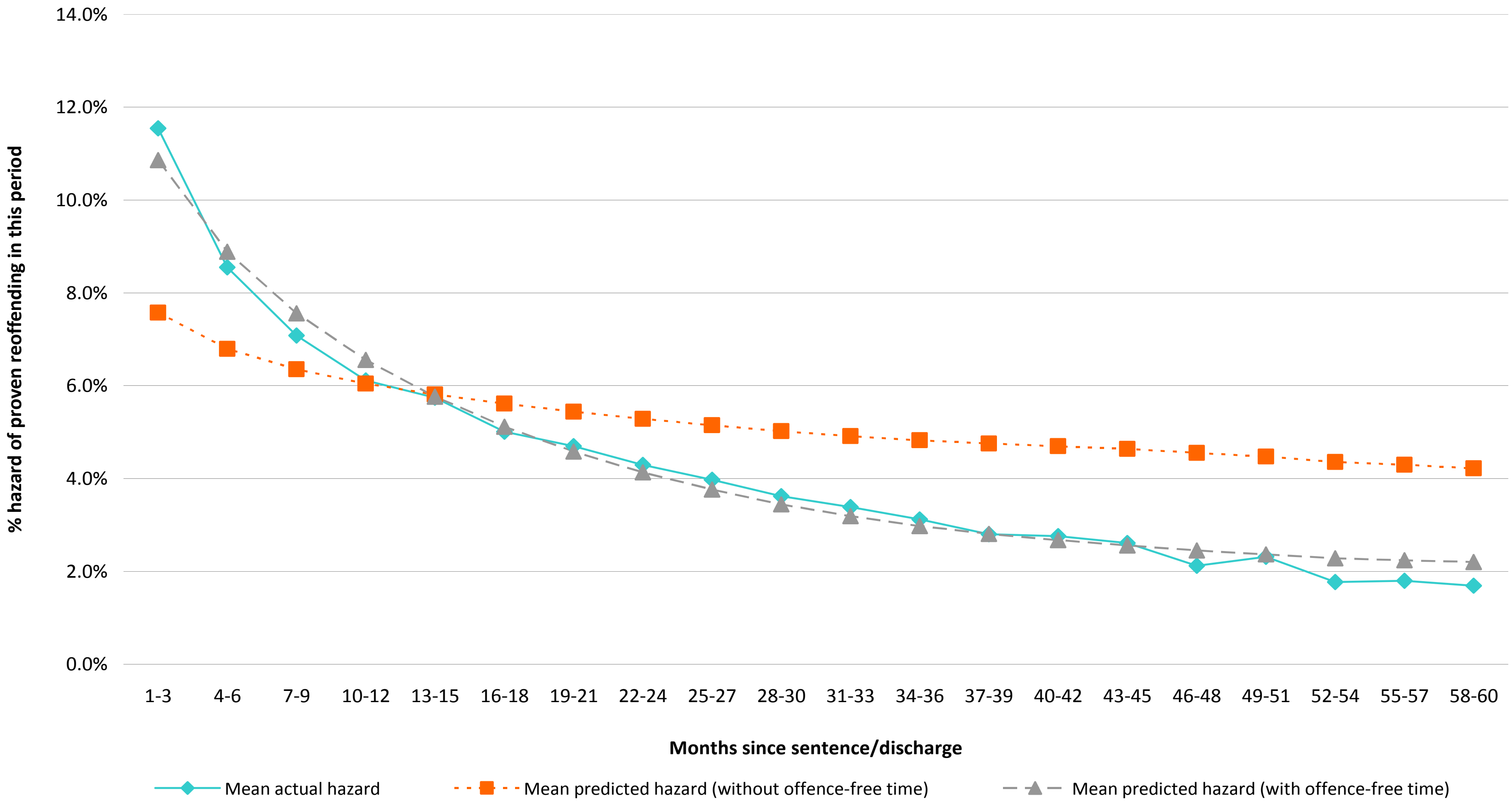
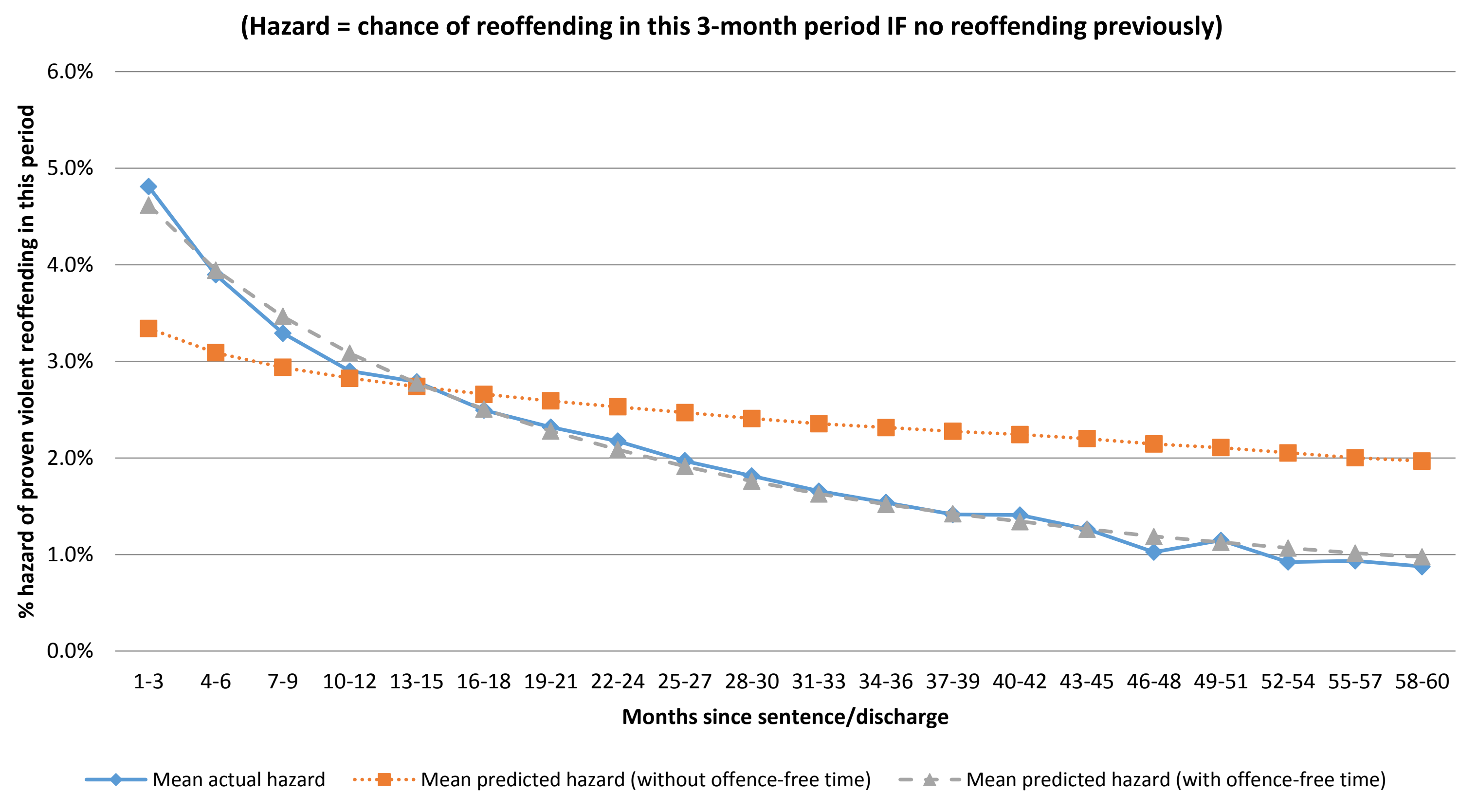


Figure 9.2: Prediction of proven violent reoffending by OVP1 score and offence-free time

(Hazard = chance of reoffending in this 3-month period IF no reoffending previously)



Optimising the use of dynamic risk factors to predict reoffending

Table 9.3 presents the associations with general and violent proven reoffending (for the construction sample) across a range of dynamic risk factors.[153] General reoffending rates were highest among drug users, peaking at 62% for heroin users and above 50% for users of most other drugs.[154] Daily drug users and those not well motivated to tackle drug misuse had similarly high rates, although these factors were highly correlated with drug type. Those with strongly pro-criminal attitudes, and whose regular activities strongly encourage offending, also had rates above 55%. Partner-related issues (3.2, 6.4 and 6.7) and motivation to address offending appeared to be weakly associated with reoffending. Some of these factors may however have predictive value due to their high frequency: for example, 28% of offenders lived with a partner compared with 11% being of no fixed abode, and 60% scored above zero for motivation compared with 25% for pro-criminal attitudes.[155] Cognitions related to the current offence (2.6 and 2.11) and emotional well-being factors were unassociated with general reoffending.

The majority of risk factors had similar associations with general and violent reoffending. However, the cocaine-based drugs were most strongly associated with violent reoffending, with methadone in particular having a weak association. Both types of alcohol misuse (chronic and binge) were strongly associated with violent reoffending (rates around 30%) but moderately associated with general reoffending. Similar patterns were found for aggressive/controlling behaviour and temper control, with the latter being somewhat more predictive of both outcomes.

Table 9.3: Associations between dynamic risk factors and proven reoffending

Risk factor		General reoffending % by risk factor score (overall rate 33.9%)			Violent reoffending % by risk factor score (overall rate 19.6%)		
		0 (No problems)	1 (Some problems)	2 (Sig. problems)	0 (No problems)	1 (Some problems)	2 (Sig. problems)
2.6	Acknowledges impact of offending	33	n/a	37	22	n/a	19
2.11	Accepts responsibility for offending	34	n/a	32	20	n/a	20
3.2	Lives with partner [no=problem]	30	n/a	36	17	n/a	21
3.3	No fixed abode	33	n/a	46	19	n/a	27
3.4	Suitability of accommodation	30	41	45	17	24	27
3.6	Suitability of accommodation location	30	42	45	17	25	27
4.2	Currently unemployed	27	n/a	41	16	n/a	24
4.7	Reading/writing/numeracy problems	31	42	44	18	25	27

153 It is not possible to meaningfully present correlations, as the prevalence of risk factors varies so greatly. Correlation coefficients are very low for rare risk factors, even when they are strongly associated with reoffending. Given the size of the sample, significance levels are also not given, as almost all associations are highly statistically significant even where risk factors are weakly associated with reoffending.

154 Information on individual drug use was often incomplete in the early years of OASys use, and these are therefore not scored in OGP1/OVP1, but is now available for almost all drug misusing offenders.

155 A predictive risk factor which is present for around 50% of offenders will offer the greatest help in efficiently separating higher and lower risk offenders. This must be balanced against effect size: a rare but potent risk factor may offer similar overall discriminatory power to a frequent but weakly associated risk factor.

Risk factor		General reoffending % by risk factor score (overall rate 33.9%)			Violent reoffending % by risk factor score (overall rate 19.6%)		
		0 (No problems)	1 (Some problems)	2 (Sig. problems)	0 (No problems)	1 (Some problems)	2 (Sig. problems)
5.2	Current financial situation	28	38	41	17	22	23
5.3	Financial management problems	29	39	43	17	22	24
5.4	Illegal activities a major income source	31	47	51	19	23	24
5.5	Overreliance on others for finance	30	42	47	17	24	27
5.6	Severe impediment to budgeting	31	42	46	18	23	25
6.1	Current relationship with family	30	39	39	17	23	25
6.4	Current relationship with partner	33	36	35	19	21	23
6.7	Perpetrator of domestic violence	33	n/a	37	18	n/a	25
7.2	Regular activities encourage offending	26	43	55	15	26	32
8.1: Main drug	Heroin	33	n/a	62	19	n/a	23
	Methadone	34	n/a	54	19	n/a	21
	Crack cocaine	34	n/a	57	19	n/a	33
	Cocaine hydrochloride [i.e. powdered]	34	n/a	51	19	n/a	35
	Amphetamines	34	n/a	55	20	n/a	29
	Cannabis	32	n/a	49	18	n/a	30
	Any other drug	34	n/a	52	20	n/a	30
Usage	[Daily drug use =2, weekly=1, else 0]	30	52	56	18	29	29
8.8	Motivation to tackle drug misuse	28	52	50	18	29	29
9.1	Current [chronic] alcohol misuse	31	36	43	16	24	31
9.2	Current binge drinking	31	35	41	15	23	29
10.2	Current psychological problems	34	35	35	19	20	21
10.6	Current psychiatric problems	34	35	35	19	22	23
10.7	Current/pending psychiatric treatment	34	n/a	33	19	n/a	22
11.2	Impulsivity	23	36	49	12	21	30
11.3	Aggressive/controlling behaviour	29	39	41	13	25	30
11.4	Temper control	28	38	44	13	25	32
11.6	Problem solving skills	23	35	44	13	20	27
11.7	Awareness of consequences	26	36	42	14	21	25
11.9	Understands others views	29	40	41	15	24	27
12.1	Pro-criminal attitudes	29	48	59	17	27	35
12.8	Motivated to address offending	27	39	38	15	23	22

Note. For items 2.6 and 2.11, the risk factor is present if the description is not applicable (e.g. does not acknowledge impact). Items 2.6, 2.11, 3.2 , 6.7 and the main drug items do not have 0/2 scoring on the OASys form: rather, 0 is used here to indicate the risk factor is not present and 2 to indicate that it is present.

Potential interactions between dynamic risk factors were generally ignored: there are many such interactions, which creates statistical power problems, and few clear theoretical cases were presented for their existence. One potential interaction was taken forward, relating to item 6.4 (Current relationship with partner, or satisfaction with single state). OGP1 and OVP1 omitted this item, but its failure to aid prediction when OGP/OVP1 were produced may have been because of its complex definition. During the years covered by our assessment sample, no direct information on whether the offender was single rather than in a relationship was available.[156] However, item 3.2 (Who does the offender live with?) includes a 'Partner' checkbox, and the interaction between 6.4 and this checkbox was therefore used to test whether the impact of relationship satisfaction varied according to whether the offender currently lived with a partner.

Indeed, some evidence of an interaction was found. Overall reoffending rates ranged from 27% (score zero on question 6.4) to 36% (score 2 on question 6.4) among those living with a partner, but did not vary by question 6.4 (35% for scores of both 0 and 2) among those not living with a partner. Violent reoffences displayed a similar pattern, ranging from 15% to 24% among those living with a partner, and from 20% to 22% for those not doing so. The OGP2 and OVP2 scoring systems therefore include interactions whereby the respectively positive and risk-raising effects of scores of 0 and 2 on question 6.4 are magnified for offenders living with a partner.

Constructing statistical models to predict all reoffending, and non-sexual violent reoffending

Exploratory analysis on the survival dataset revealed that little advantage would be being gained from using variables or Copas formulations which differed from those in OGRS4 (see Chapter 8). A decision was therefore taken to limit static risk factors to those included in the OGRS4 equations, and select the same Copas rates.[157]

Updating the user-friendly OGP and OVP risk predictors

Figures 9.3 and 9.4 compare the absolute predictive validity of differing model types. In both figures, residuals are smoothed for clarity.[158] Standard 'with-OGRS' models, which simply used a proportion of OGRS4 scores to provide their static/O-FT element, showed strong residual patterns. For both types of reoffending, there was under-prediction (positive residuals) across a broad range from low-average to high-average risk but over-prediction (negative residuals) for the very lowest and very highest risk offenders. The consequences of this appeared fairly serious around the risk levels where the pattern

---

156 The full version of layered OASys, implemented in August 2009, includes question 6.8: Current relationship status. As implementation post-dated our sampling period, it was not possible to use this item in the current study. Its use could be trialled in future studies.

157 Predictive validity gains from using alternative indicator variables or Copas codings were very modest. These gains were judged to be outweighed by the practical benefits from using the same variables in OGP/OVP and OGRS. NOMS IT developers need to code all predictors and thus their constituent variables in multiple formats (e.g. Excel, OASys IT), before subjecting the predictors to extensive user acceptance testing to ensure consistent results. Reducing the total number of variables, especially variant codings of similar items, should aid the testing process. Similar benefits would be accrued by anyone using the predictors outside NOMS (e.g. academic researchers, other jurisdictions).

158 A simple 7-point smoothing technique was used. For example, the plotted residual for those with predictions of 15% was actually the average of all residuals for predictions between 12% and 18%. Before this occurred, interpolation was used to estimate residuals for "missing" predicted percentages. For example, an offender scoring 55/100 on the no-OGRS OGP2 risk factors would have a next-two-year prediction of 51%, whereas one scoring 56/100 would have a prediction of 53%. As predictions of 52% therefore will never occur, the 52% residual was estimated as a simple average of the 51% and 53% residuals. Finally, smoothed residuals were presented where the number of cases in the validation sample averaged at least 100 for each of the seven residuals to be averaged (and then only when this criterion was met for each of the three model variants).

switched. For example, as predicted likelihood of general reoffending rose from 49% to 56%, the actual rate only rose by around 2%. The problem is also evident for violent reoffending, where actual reoffending rates for offenders only rose by 5% as predictions rose from 31% to 42%.

Because of this clear failing of the standard ‘with-OGRS’ models, alternate versions were trialled, using a quadratic transformation of the OGRS scores. (See Appendix I for a summary of the models). These sacrificed no predictive validity,[159] and in the case of OVP2 resulted in much smaller residuals than the standard models, though the ‘no-OGRS’ models produced still smaller residuals. The quadratic models sacrificed the practitioner benefit of relating the static/O-FT element to the OGRS score in a simple manner, but were still easier than the ‘no-OGRS’ models from the perspective of those coding scores (i.e. IT developers, researchers). On this basis, the no-OGRS and quadratic with-OGRS models were taken forward.

Table 9.4 compares the absolute and relative predictive validity of the no-OGRS and quadratic with-OGRS models, for offenders of varying age, gender, ethnicity and sentence. The no-OGRS predictors had smaller residuals in some general recidivism comparisons, with little difference in net performance across the other general recidivism or any violent recidivism comparisons. Specifically, quadratic with-OGRS fared worse in that it under-predicted female and 18–20 year olds general recidivism, and over-predicted general recidivism in the over-50s. The most notable results common to both predictor types are perhaps related to offenders discharged from short custodial sentences. This group’s general reoffending rate was almost 6% greater than predicted, and their violent reoffending rate was almost 4% greater than predicted. In contrast, the small number of offenders discharged from indeterminate sentences reoffended at approximately one-third of their already low predicted rates. Both groups may be affected by selection issues – few short custodial sentenced offenders are assessed, as only young adults are supervised after release from a short sentence, while the Parole Board may release only those indeterminate sentenced prisoners who can be managed safely in the community. In addition, those indeterminate sentenced prisoners who are released may be subject to stringent risk management in the community. Both predictor types also under-predicted general recidivism by Asian and Black offenders, and under-predicted violent recidivism by Black and Mixed ethnicity offenders.

For both reoffending outcomes, the relative predictive validity of the no-OGRS predictor was fractionally better overall. For most sub-groups, the no-OGRS predictors had AUCs which were one- to five-thousandths better than those of the with-OGRS predictors. More variable results were found for the smallest sub-groups of offenders.

Relative predictive validity should usually be lower for age and gender sub-groups than overall, as these characteristics are scored in the predictors and heterogeneity within sub-groups is therefore reduced. In fact, AUCs were above average for female offenders and, for violent reoffending, older

[159] In fact, AUC values for the quadratic models were .001 or .002 greater than for the standard with-OGRS models.

offenders. AUCs were lower for the youngest two age groups, where criminal histories were less diverse. AUCs were also lower for some Black, Asian and Minority Ethnic (BME) groups, especially Black offenders, and they were lower for offenders given non-custodial or long custodial sentences, compared with those given suspended, short- or medium-length custodial sentences. The AUCs for indeterminate and unknown length sentences have broad confidence intervals due to their small sample sizes.

Table 9.4: Actual and predicted general and violent reoffending for with-OGRS and no-OGRS models

	General reoffending					Violent reoffending				
Offender group ***(number / % of offenders)***	**Actual**	**Residual for with-OGRS**	**Residual for no-OGRS**	**AUC: with-OGRS**	**AUC: no-OGRS**	**Actual**	**Residual for with-OGRS**	**Residual for no-OGRS**	**AUC: with-OGRS**	**AUC: no-OGRS**
All offenders										
(validation only: 36,221)	33.5%	-0.2%	+0.1%	0.775	0.776	19.8%	+0.3%	+0.3%	0.763	0.764
(all: 109,319)	33.8%	-0.1%	+0.1%	0.776	0.778	19.7%	+0.1%	+0.1%	0.766	0.768
Gender										
Female *(5,320 / 14.7%)*	24.9%	-1.4%	-0.6%	0.788	0.790	13.1%	-0.3%	+0.6%	0.784	0.783
Male *(30,901 / 85.3%)*	35.0%	Zero	Zero	0.770	0.772	20.9%	+0.4%	+0.2%	0.756	0.757
Ethnicity										
Asian *(4,258 / 3.9%)*	32.0%	+3.7%	+3.3%	0.759	0.761	15.5%	+0.6%	+0.3%	0.752	0.757
Black *(5,139 / 4.7%)*	37.9%	+4.1%	+4.3%	0.751	0.754	19.9%	+1.9%	+2.2%	0.739	0.742
Mixed *(2,029 / 1.9%)*	41.5%	+1.3%	+1.5%	0.763	0.764	25.0%	+1.6%	+1.7%	0.757	0.758
Other *(635 / 0.6%)*	25.4%	+0.9%	-1.6%	0.774	0.769	12.3%	-0.2%	-0.4%	0.802	0.807
White *(83,562 / 76.4%)*	34.8%	-0.4%	-0.3%	0.779	0.780	20.5%	-0.1%	Zero	0.767	0.768
Missing / not stated *(13,696 / 12.5%)*	25.8%	-0.9%	-1.0%	0.757	0.759	15.5%	-0.5%	-0.5%	0.764	0.768
Age										
18–20 *(5,356 / 14.8%)*	46.5%	+1.1%	Zero	0.745	0.746	31.0%	+0.4%	-0.7%	0.728	0.728
21–24 *(6,136 / 16.9%)*	39.9%	-0.2%	-0.6%	0.741	0.743	24.9%	Zero	-0.4%	0.731	0.733
25–29 *(6,165 / 17.0%)*	37.8%	+0.3%	+0.3%	0.757	0.759	21.1%	+0.5%	+0.3%	0.736	0.737
30–34 *(5,036 / 13.9%)*	33.8%	-0.3%	-0.1%	0.759	0.759	18.7%	+1.3%	+1.0%	0.739	0.736
35–39 *(4,697 / 13.0%)*	29.0%	-0.6%	-0.2%	0.772	0.772	16.1%	-0.7%	-0.6%	0.740	0.744
40–49 *(6,132 / 16.9%)*	24.2%	-0.4%	+0.4%	0.758	0.760	13.5%	+0.2%	+0.9%	0.752	0.753
50–59 *(2,028 / 5.6%)*	12.8%	-3.3%	-1.2%	0.747	0.755	6.0%	-2.5%	-0.4%	0.792	0.810
60+ *(671 / 1.9%)*	8.5%	-2.0%	+0.8%	0.773	0.762	3.9%	-1.4%	+1.2%	0.846	0.851

Offender group *(number / % of offenders)*	**General reoffending**					**Violent reoffending**				
	Actual	**Residual for with-OGRS**	**Residual for no-OGRS**	**AUC: with-OGRS**	**AUC: no-OGRS**	**Actual**	**Residual for with-OGRS**	**Residual for no-OGRS**	**AUC: with-OGRS**	**AUC: no-OGRS**
Sentence type and length										
Non-custodial *(63,794 / 58.4%)*	31.0%	-0.4%	-0.5%	0.760	0.762	18.4%	Zero	-0.3%	0.756	0.758
Suspended *(19,592 / 17.9%)*	35.1%	-0.3%	-0.5%	0.777	0.779	20.5%	-0.4%	-0.7%	0.766	0.769
Custodial: short *(7,411 / 6.8%)*	56.4%	+5.7%	+5.7%	0.780	0.782	34.4%	+3.8%	+3.6%	0.744	0.744
Custodial: medium *(12,743 / 11.7%)*	35.1%	-0.8%	+0.3%	0.791	0.793	18.8%	-0.4%	+0.5%	0.777	0.782
Custodial: long *(5,400 / 4.9%)*	29.0%	-0.7%	+1.1%	0.762	0.764	14.0%	-0.5%	+1.2%	0.758	0.766
Custodial: indeterminate *(264 / 0.2%)*	7.6%	-13.0%	-10.8%	0.756	0.763	4.2%	-7.9%	-7.2%	0.807	0.801
Custodial: length unknown *(116 / 0.1%)*	36.2%	-2.1%	-1.7%	0.821	0.819	19.8%	-1.0%	-1.1%	0.812	0.832

Note. Validation sample used for age and gender analysis. Entire sample used for ethnicity and sentence analysis.

Figure 9.3: Absolute predictive validity (actual minus predicted general reoffending residuals) of three methods for modelling static/offence-free time risk in OGP2

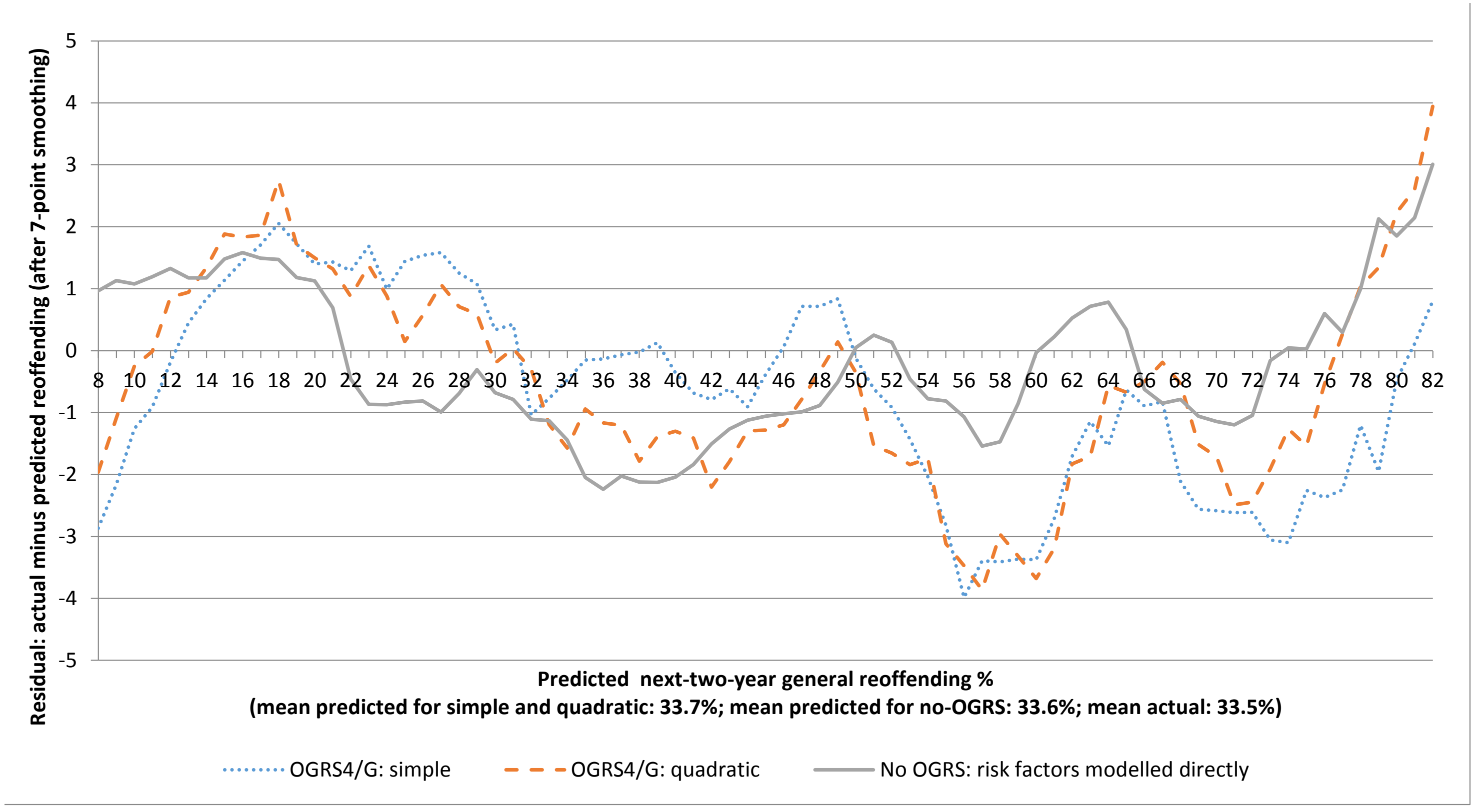
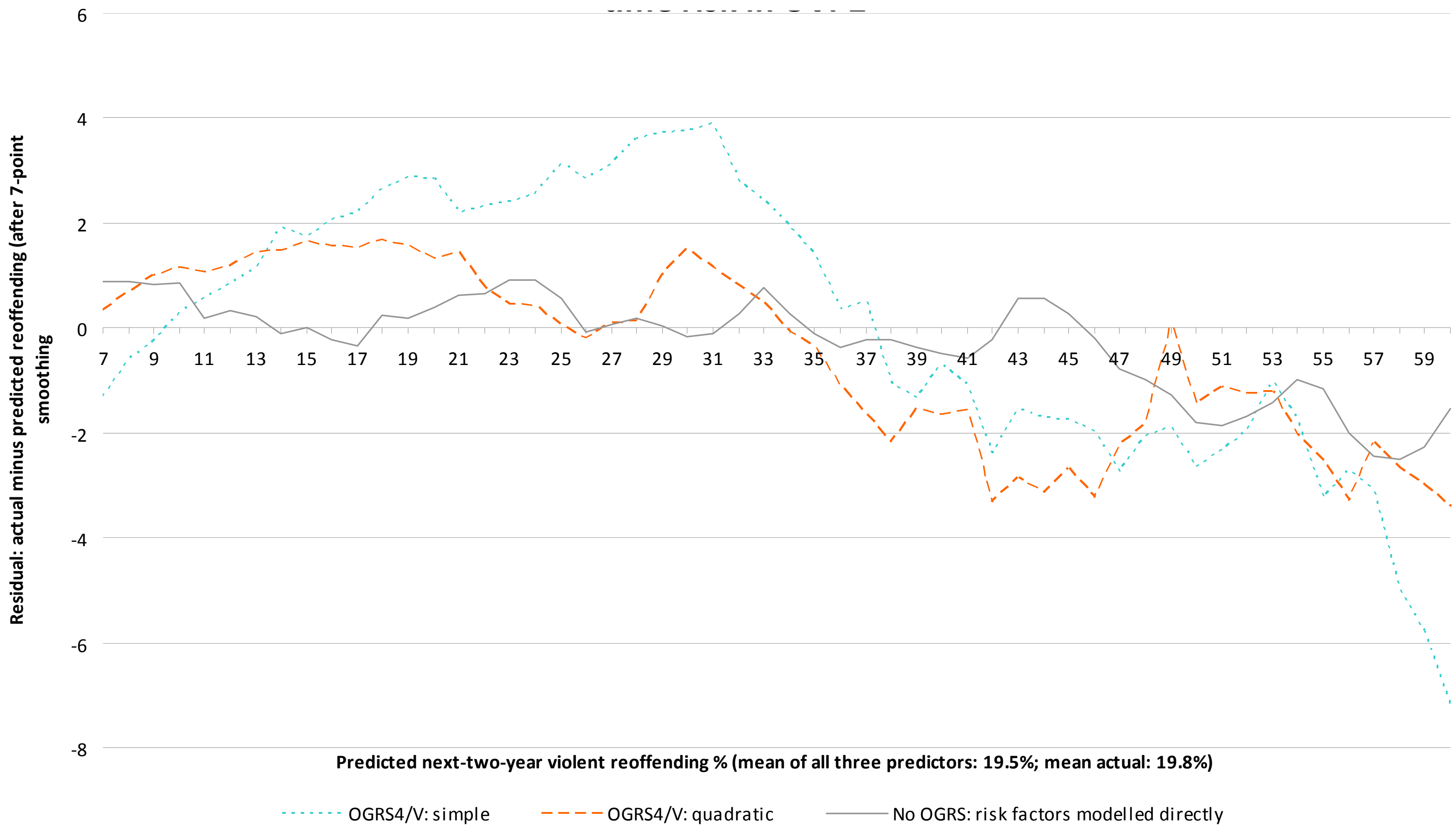


Figure 9.4: Absolute predictive validity (actual minus predicted violent reoffending residuals) of three methods for modelling static/offence-free time risk in OVP2



Validating the new OGP2 and OVP2 risk predictors

Table 9.5 demonstrates that for all reoffending, OGP1 was outperformed by both versions of OGP2 and also by OGRS4/G. There was little difference in the overall AUCs of the OGP2 versions, with the no-OGRS version being fractionally but significantly better (by 0.0016 points of AUC). For violent reoffending, there were no statistical difference between the with-OGRS and no-OGRS versions of OVP2 (*p*=.16), but the results confirmed that violence-specific predictors incrementally improved upon general predictors.[160]

Table 9.5: Predictive validity comparisons of OGRS4, OGP1/OVP1, and with-OGRS and no-OGRS models for OGP2/OVP2

Predictor	AUC (95% confidence limits) by reoffending outcome	
	All offences	**Violent offences**
Predictors of all reoffending		
OGRS4/G	.766 (.761, .771) ***	.737 (.730, .744) ***
OGP1	.755 (.749, .760) ***	n/a
OGP2: with-OGRS (quadratic)	.775 (.770, .780) **	.745 (.739, .751) ***
OGP2: no-OGRS	.776 (.771, .781) [RC]	.746 (.740, .752) ***
Predictors of violent reoffending		
RM2000/v	n/a	.689 (.681, .694) ***
OGRS4/V	n/a	.752 (.746, .758) ***
OVP1	n/a	.746 (.740, .752) ***
OVP2: with-OGRS (quadratic)	n/a	.763 (.757, .769)
OVP2: no-OGRS	n/a	.764 (.758, .770) [RC]

Note. *N* = 36,621; *n* / % of reoffenders = 12,128 / 33.5% (all), 7,157 / 19.5% (violent). RC = reference category.
*: *p*<.05. **: *p*<.01. ***: *p*<.001.

Table 9.6 repeats these comparisons for offenders starting their community sentence or being discharged from custody in March 2008 (i.e. where all offenders had zero offence-free months). This comparison is very important in terms of the validity of the new predictors for court reports, assessments completed prior to discharge, and assessments completed at the start of the probation period. It therefore checks that OGP2 and OVP2 are, at least, no worse than their predecessors in these key situations.

The AUCs in Table 9.6 were higher than those in Table 9.5 for the most successful predictors of reoffending. This reflects the greater heterogeneity and therefore ease of prediction in the zero O-FT caseload compared with the overall caseload, as the latter group contained fewer offenders with high likelihoods of reoffending (because more of the higher risk cases reoffended prior to their potential follow-up start date). There were no significant differences between the with-OGRS and no-OGRS

160 When restricting the outcome measure to homicide/wounding reoffences, the confidence intervals were wide, with the only definite result being that RM2000/v was a markedly inferior predictor.

versions of OVP2, but the no-OGRS version of OGP2 predicted general reoffending with significantly greater success.

Table 9.6: Predictive validity comparisons: cases with zero offence-free months

Predictor	AUC (95% confidence limits) by reoffending outcome	
	All offences	**Violent offences**
Predictors of all reoffending		
OGRS4/G	.772 (.749, .796) ***	.727 (.701, .753) ***
OGP1	.779 (.755, .802) ***	n/a
OGP2: with-OGRS	.795 (.773, .817) **	.743 (.718, .769) ***
OGP2: no-OGRS	.801 (.779, .823) [RC]	.746 (.721, .772) ***
Predictors of violent reoffending		
RM2000/v	n/a	.685 (.658, .713) ***
OGRS4/V	n/a	.759 (.733, .784) ***
OVP1	n/a	.762 (.737, .787) ***
OVP2: with-OGRS	n/a	.782 (.758, .806)
OVP2: no-OGRS	n/a	.786 (.762, .810) [RC]

Note. N = 1,501; n of reoffenders = 793 (all), 494 (violent). RC = reference category. *: p<.05. **: p<.01. ***: p<.001.

Based on the results in Tables 9.5 and 9.6, it was confirmed that OGP/OVP2 improved prediction compared with OGP/OVP1. Based on Tables 9.4, 9.5 and 9.6, it was determined that the no-OGRS versions offered better absolute predictive validity for women offenders and the oldest and youngest offenders, as well as marginally better relative predictive validity across the whole caseload. Therefore, the with-OGRS versions were set aside, and the no-OGRS versions were taken forward as the final versions of OGP2 and OVP2.

Appendix I includes the results of the initial logistic regression models on the next-two-year construction sample. Table 9.7 presents the final OGP2 and OVP2 scoring tables, supplemented by full details of offence-free time scoring in Table 9.8. In OGP2, it proved possible to allocate 15 points to offence-free time and 35 points to dynamic risk factors, with the remaining 50 points for static risk factors. In OVP2, these allocations were 10, 30 and 60 points respectively. These risk factor weightings were found to assist with simplicity, allowing many 0/2 or 0/1/2 items to have two or four point weightings. The process which created the 100-point scores from the initial logistic regression models is further explained in Appendix I (which also gives full details of the static scoring system). For OGP2, there was around a 16-point difference in the average scores of reoffenders and non-reoffenders. For OVP2, the difference ranged from 10 to 15 points depending on the measure employed.[161] For the selected OGP2 and OVP2 scoring rules, adjustments in the OGP2 score (so that it was not lower than the OVP2 score) had to be made for 0.8% of cases. The mean adjustment was 2.5%. The majority of adjustments were of 1% or 2%, but adjustments of at least 5% were required for about one-eighth of affected cases (i.e. 0.1% of all cases).[162]

Comparison with Table 9.1 shows several differences in dynamic risk factors between versions 1 and 2 of OGP/OVP. The ‘partner score’, consisting of three items relating to intimate partner relationship and living arrangements, is new to both predictors yet now has a substantial weighting in both, suggesting that the reliability and quality of such items improved between the 2002–04 version 1 construction period and the 2005–08 version 2 construction period.

The introduction of individual drugs has substantially changed the predictors: in OGP, it edges out all previously scored drug misuse items other than frequency of use, while it allows drug misuse items into OVP for the first time. Differences between opiates and cocaine-based drugs are evident: the former are only related to non-violent reoffending, while the latter are good predictors of violent reoffending.[163] Together with the frequency of use item, this illustrates the importance in general reoffending of acquisitive behaviour as a result of drug addiction, whereas OVP features the use of drugs which trigger angry or aggressive behaviour (Boles and Miotto, 2003).

The shift of OGP’s focus from non-violent to any reoffending allows alcohol misuse and temper control, which are important violence risk factors, to be included in OGP as well as OVP, though with smaller weightings. Version 1 included several thinking and behaviour items in OGP but only temper control in OVP; in version 2, different single items accompany temper control in both versions.

161 For those with no reoffending, OGP2 100-point scores had a mean of 37.9 points, median of 38.0 points and a mode of 41 points (suggesting some skewness in the distribution). For those with any reoffending, these were 53.3, 54 and 57 points respectively. The standard deviations for these OGP2 scores were 14.3 and 13.9 points for non-reoffenders and reoffenders respectively.. For those with no violent reoffending, OVP2 100-point scores had a mean of 46.7 points, median of 48 points and mode of 51 points (again, suggesting some skewness), while for violent reoffenders these were 61.4, 62, and 61 points respectively. Overall means and standard deviations were 43.1 and 16.0 for OGP2 and 49.6 and 15.9 for OVP2

162 An alternative approach was considered, which would have entirely avoided the problem of inconsistent results. This would require the OVP2 probability to equal the product of the OGP2 probability and the conditional probability that any reoffending would include some violent reoffending. The complexity of this approach, from the perspective of assessors, is considerable. (In considering an offender’s dynamic risk factors for violence, they would have to consider the factors in both OGP2 and in the latter conditional probability equation, which would necessarily be different at least in weighting and perhaps in nature.) Therefore, given the low prevalence of inconsistent results and the small difference in the two probabilities when inconsistency did occur, the pragmatic solution was followed.

163 Logistic regression models have also been used to analyse the Surveying Prisoner Crime Reduction (SPCR) prisoner cohort data, examining the associations with reoffending of various offender, offence and sentence attributes. The dynamic risk factors found to be associated with increased odds of reoffending within one year were: (i) homeless or living in temporary accommodation prior to sentence; (ii) use of Class A drugs since release from prison; and (iii) having regularly playing truant from school (Brunton-Smith and Hopkins, 2013).

Table 9.7: OGP2 and OVP2 scoring

Risk factor	Max raw score or categories	Weight in OGP2	Weight in OVP2
OGP2 static score	See note	50	--
OVP2 static score	See note	--	60
Total static score		***50***	***60***
Offence-free time	See Table 9.8	15	10
Total offence-free time score		***15***	***10***
3.4 Suitability of accommodation	2	2	2
4.2 Employment status	2	2	2
3.2, 6.4 and 6.7: Partner score	4: see note	4	4
7.2 Regular activities encourage offending	2	4	--
8.1 Main drug used	Crack Amphetamine Heroin Cocaine Cannabis Methadone Any other None	7 6 5 4 1 0 2 0	6 3 0 4 2 0 1 0
8.4 Frequency of main drug use	Daily Weekly Less frequent	2 1 0	--
9.1 and 9.2 Current alcohol use and binge drinking	2+2	2+2	4+4
11.2 Impulsivity	2	2	2
11.4 Temper control	2	2	4
11.6 Problem solving skills	2	4	2
12.1 Pro-criminal attitudes	2	2	--
Total dynamic score		***35***	***30***
Total score		***100***	***100***
Logistic function parameter	Intercept	-4.1376 (0.0337)	-5.5545 (.0491)
(standard error)	Beta	0.0744 (0.000670)	0.0760 (.000826)

Note. 'Partner score' is calculated as the 6.4 score (on a 0/1/2 scale), plus 1 point if not living with partner (item 3.2), plus 1 point if a domestic violence perpetrator (item 6.7). The process of calculating 50- and 60-point static scores is described in Appendix I. In the logistic function, z = intercept + (Beta*total score), next-two-year probability = $e^z / (1 + e^z)$, and next-two-year percentage is 100 * probability, rounded to the nearest whole number. If necessary, raise the OGP2 next-two-year percentage to equal the OVP2 next-two-year percentage.

Table 9.8: Scoring of offence-free time in OGP2 and OVP2

Complete months since index date	OGP2 points	OVP2 points
0	15	10
1	14	10
2	14	9
3, 4	13	9
5, 6	12	8
7	11	8
8	11	7
9, 10	10	7
11, 12	9	6
13	8	6
14	8	5
15, 16	7	5
17	6	4
18, 19, 20	5	4
21, 22	5	3
23, 24	4	3
25	4	2
26, 27, 28	3	2
29, 30, 31	2	1
32, 33	1	1
34	1	0
Above 34	0	0

Two risk factors, both of which were only in OVP1, are entirely eliminated. Acknowledging the impact of offending and current psychiatric treatment were both found to be quasi-static in the study of change in OGP and OVP scores (see Chapter 4), in that both changed infrequently and those changes which did occur were not predictive of later violent reoffending. Other risk factor domains are changed in a more subtle manner, if at all. For accommodation, the suitability item is preferred to the no fixed abode item. Suitability can be scored either 0, 1 or 2 when the offender has a fixed abode, though is always scored 2 when the offender is of no fixed abode. As in version 1, current employment status features in both predictors while activities which encourage offending are in OGP only.

Validating OGP2 and OVP2 risk bands

Table 9.9's sensitivity and specificity statistics present predictors' performance in an alternate way to the AUC. While the AUC has summary utility, as the probability that a randomly selected reoffender will have a higher score than a randomly selected non-reoffender, sensitivity and specificity have practical value as they focus upon the boundaries between risk categories. The results suggest that OGRS4/G and OGP2 bandings provide considerable improvement upon both OGRS3 and OGP1. OGP1 performed only slightly better than OGRS3 at placing reoffenders into the high and very high risk groups, whereas OGRS4/G did much better and OGP2 did better still for all three outcomes. In the prediction of violent reoffending, at least half of the potential improvement in each of the three true positive counts comes from switching from OGRS3 to OVP1 – that is, from a general reoffending predictor to the weakest of the three violence-specific predictors. At the lowest cutoff, the majority of the remaining improvement comes from the improvements specific to OVP2, rather than OGRS4/V.

For all reoffending, it can be seen that 43% of offenders were at least medium risk; this 43% included 67% of reoffenders when OGRS3 was used, rising to 70% when OGP2 was used, and also included over 31% of non-reoffenders under OGRS3 but under 30% under OGP2. The 11% who were at least high risk included between 22.0% and 24.3% of reoffenders, while the 1.3% who were very high risk included between 3.1% and 3.6% of reoffenders. Among violent reoffending, the dramatic improvement associated with moving to a violence-specific predictor can be seen, as the 4.2% of offenders in the high or very high group included 9.7% of reoffenders under OGRS3 but 12.3% under OVP1, improving further to 12.7% under OGRS4/V and 13.2% under OVP2.

Table 9.9: Sensitivity and specificity of risk predictor categories

Risk categories / score range below cutoff	Risk categories / score range above cutoff	% of all offenders above cutoff	Predictor	Sensitivity	Specificity
All offences (33.5% proven reoffending)					
Low	Medium, high, very high	43.3	OGRS3 OGP1 OGRS4/G OGP2	67.1 68.5 69.4 70.4	68.8 69.5 69.9 70.4
Low, medium	High, very high	10.7	OGRS3 OGP1 OGRS4/G OGP2	22.0 22.4 23.9 24.3	95.1 95.3 96.0 96.2
Low, medium, high	Very high	1.3	OGRS3 OGP1 OGRS4/G OGP2	3.1 3.3 3.4 3.6	99.6 99.6 99.8 99.9
Violent offences (19.8% proven reoffending)					
Low	Medium, high, very high	34.2	OGRS3 OVP1 OGRS4/V OVP2	57.7 62.2 63.1 64.9	71.6 72.7 72.9 73.4
Low, medium	High, very high	4.2	OVP1 OGRS3 OGRS4/V OVP2	9.7 12.3 12.7 13.2	97.2 97.9 98.0 98.1
Low, medium, high	Very high	0.3	OVP1 OGRS3 OGRS4/V OVP2	0.9 1.2 1.4 1.4	99.8 99.9 99.9 99.9

Note. For the purpose of comparison, score ranges were set to match the distribution of OGP1 (all reoffending) or OVP1 (violent reoffending) categories, with ties broken randomly. For OGRS3, these are not the score ranges used in NOMS practice; for OGRS4/G and V, OGP2 and OVP2, they do not represent proposed score ranges. ‘Per 10,000 cases’ results represent a 10,000-strong caseload with identical reoffending rates and score distributions to the original 36,221 cases. Due to rounding, some rows do not add to 10,000. Sensitivity = TP/(TP+FN), i.e. of reoffenders, the proportion predicted to reoffend. Specificity = TN/(TN+FP), i.e. of non-reoffenders, the proportion predicted to not reoffend. TP = true positive; FN = false negative; FP = false positive; TN = true negative.

9.4 Implications

The results in this chapter demonstrate significant benefits in updating OASys's predictors of reoffending from version 1 to version 2 of OGP and OVP. Relative predictive validity increases by between 1 and 2 points of AUC for each type of reoffending, providing improved ability to distinguish higher from lower risk offenders when determining RoSH (for OVP), resource levels and suitability for interventions, and when reporting prior to sentence and discharge from custody (e.g. parole, Home Detention Curfew). These AUC improvements are pushing back the threshold of what is achievable, given that OGP/OVP1 and OGRS4 have already been developed on the basis of effective yet user-friendly prediction methodology. While methodological changes make the static risk scores more of a 'black box' than previously, they have allowed actual and predicted rates to match closely for offenders of all ages and both genders, resolving a problem which Chapter 3 had noted in OGP/OVP version 1.

Any decision to change the predictors used in practice would need to weigh these benefits against the costs of implementation. These would include revisions to the OASys IT system and the production and dissemination of assessor guidance. Given that the implementation of OGP/OVP version 1 did not require specific assessor training, it can be assumed that version 2 would also not require this resource-intensive activity. The greatest innovation in version 2 is the introduction of offence-free time as a risk factor. Assessor guidance would need to carefully explain the next-two-year concept, and the consequential fall in offenders' scores as they remain offence-free. Assessors would also need to understand the requirement to rescore offenders (with increased sanction counts and offence-free time set to zero) if proven reoffending occurs, even if no new custodial sentence or court order is imposed.

Offence-free time increases the existing benefits of reviewing OASys assessments periodically. As shown in Chapter 4, the majority of dynamic version 1 score changes are downwards. Given the substantial overlap between the two versions' dynamic items, the same will be true of version 2, but with the additional and substantial effects of offence-free time. Importantly, reductions in scores could be used to justify reductions in staff contact with these offenders, enabling the reallocation of resources to other higher risk cases. As revisions to NOMS National Standards (National Offender Management Service, 2011b) allow for more discretion in the completion of assessment reviews, effective communication of these review benefits will be vital to enable managers to make informed decisions about local practice.

As Howard (2011) noted, the decline in reoffending hazards as offence-free time passes has implications for intervention targeting. Offenders who are initially eligible for specific interventions might become ineligible as time passes. If an offender appears unlikely to secure a place on an intervention for which they are only marginally eligible at the start of their order/licence, it could often be appropriate to substitute a lower-intensity form of treatment which can be delivered immediately, close to the point of their peak hazard.

If implemented in NOMS practice, the predictors should also be made available to researchers. Controlling for predicted rates of recidivism – whether through case/control matching or regression-based methods such as propensity score matching – is an important element of evaluations of rehabilitative interventions. Making allowance for offence-free time has not been undertaken in previous evaluations, and could improve the quality of future evaluations of community-based interventions where not all participants commence upon start of sentence/discharge.

Prior to implementation, further work will need to be undertaken to create the next-one-year predicted probabilities which correspond to the two-year probabilities. The changes in dynamic risk factors between versions 1 and 2 of the predictors will also require changes to the content of OASys. Item 3.2 was removed from OASys in 2009, while items 6.4, 11.2 are not currently included in standard OASys. If it is necessary to remove other currently scored questions in order to maintain the existing length of OASys, items 2.6 and 10.7 appear vulnerable in both layers (they are not required in criminogenic need scoring – see Chapter 11 and Appendix J), while items 11.7, 11.9 and 12.1 may be vulnerable in standard.

Further validation is summarised in Chapter 13. This exercise led to some change in the way the predictors are formulated, but their value as improvements over the first-generation OGP and OVP predictors and also over the new OGRS4 predictors continued to be demonstrated. Taken together, the results demonstrate that risk prediction continues to benefit from the assessment of dynamic risk factors. While the OGRS4 predictors improve upon OGP/OVP1 and OGRS3, further improvements upon OGRS4/G and OGRS4/V are made by OGP2 and OVP2. Dynamic risk assessment also allows offender managers to base sentence plans on a thorough understanding of offenders' criminogenic needs (see also Chapter 11) and responsivity issues.

10. Development of a new sexual reoffending predictor

This chapter examines whether OASys and criminal history information can be combined into a score which improves prediction of the sexual offences most likely to cause serious harm: those involving direct contact with victims. These 'contact offences', involving direct and serious harm, include rape, sexual assault, gross indecency, incest, unlawful sexual intercourse and grooming. While clearly still harmful, the sexual offences excluded from this category principally comprise those related to indecent images of children and exhibitionism (e.g. indecent exposure). The chapter sets out the following key points:

- Sexual offences were divided into four groups: contact adult, contact child, paraphilia (e.g. indecent exposure) and indecent images of children. Examining patterns of previous sanctions (i.e. cautions and convictions) and reoffending, for approximately 15,000 offenders, showed that offenders tend to strongly specialise by committing particular types of sexual offence.
- A new seven-item predictor, the OASys Sexual reoffending Predictor (OSP), was developed. This predictor uses static risk factors only and can thus be scored on the basis of summary printouts of individual offenders' demographics and criminal histories. It was found to be superior to RM2000/s as a predictor of contact sexual reoffending.
- The risk factors in OSP are (strongest first): contact adult sanctions; current age; age at last sexual offence; contact child sanctions; paraphilia sanctions; not first-time entrant; and stranger victim of current sexual offence.
- As OSP has the potential to improve prediction of those sexual offences most likely to cause serious harm, and is no more complex to administer, it is recommended that its implementation in NOMS and police practice should be considered – it has already been incorporated within the new Risk of Serious Recidivism (RSR) tool and used for segmenting the NOMS caseload. If OSP is fully implemented as a stand-alone predictor, amendments to user guidance and IT systems will be necessary.

10.1 Context

Predicting offenders' likelihood of recidivism plays an important role in determining how those convicted of sexual offences should be treated and managed. The sex offender treatment programmes run by the National Offender Management Service (NOMS) are relatively resource-intensive, while probation supervision and police monitoring also consume considerable staff time. In order to meet both services' public protection goals, these resources should be concentrated upon those convicted sex offenders most likely to commit further offences which cause serious harm.

At present, NOMS and police assess the recidivism risk of adult male sex offenders using Risk Matrix 2000/s (RM2000/s; Thornton, 2007). This tool categorises offenders into four risk bands based on their age, general and sexual criminal history, type of victim and relationship history. Treatment provided is proportionate to risk; offenders in the low risk band are directed to less intensive treatment than those in higher risk bands. They may also have less contact with their Offender Manager while supervised in the community and may be visited less frequently by specialist police staff, whereas those in the highest risk bands will be subject to more intensive supervision by probation and police.

RM2000/s is an actuarial predictor, where risk bands are scored mechanically based on fixed rules. It is supplemented by clinical ratings of risk factors specific to sexual offending through the Structured Assessment of Risk and Need (SARN; Webster *et al*., 2006, based on the Structured Risk Assessment; Thornton, 2002), though this instrument is only routinely administered for NOMS offenders who have completed treatment programmes. Psychometric tests, such as the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999), and clinical criteria are also used to determine treatment allocation. RM2000/s is nevertheless a primary determinant of treatment and supervision levels for adult male sex offenders.

The moderate predictive validity of RM2000/s has been confirmed in England and Wales (Barnett, Wakeling and Howard, 2010) and Scotland (Grubin, 2008). Furthermore, recent research (Wakeling, Mann and Milner, 2011) has identified improvements to user guidance which could help to ensure higher levels of inter-rater reliability (i.e. that an offender's risk factors and overall risk category will not vary according to who completes their risk assessment). However, recent discussions within NOMS have identified the following five concerns about RM2000/s, which suggest that it may be possible to improve the effectiveness and efficiency of the risk prediction process, and therefore make better use of the available treatment and management resources:

- Given that sexual offences range from those causing indirect (e.g. internet-related) or direct psychological harm (e.g. voyeurism, indecent exposure) to those involving intense direct physical and psychological harm (e.g. rape), it is questionable whether actuarial tools can be optimised to predict the risk of all sexual recidivism. A focus on offences causing direct and intense harm would assist the public protection aims and objectives of NOMS and the police service (Ministry of Justice, 2010b).
- RM2000/s requires the assessor to gain a relatively detailed understanding of the offender's criminal and relationship history, in order to score items on ever having a stranger victim and ever having been in a stable live-in relationship. It may be possible to make the assessment process more efficient and practical, given the limited information, staff time and expertise available in some NOMS and police assessment settings, by placing more emphasis on the offender's present situation.

- RM2000/s user guidance (Thornton, 2007) suggests that scores should not be created for those whose sexual offending all took place before age 16, leaving these offenders without any risk category.[164]
- RM2000/s does not cover all known domains of sexual recidivism risk: it is possible that the time since last sexual offence could be a useful risk factor (Hanson and Thornton, 2003). There may also be some scope to apply dynamic risk factors (Mann, Hanson and Thornton, 2010), which NOMS assessors could readily assess as part of OASys.
- RM2000/s was based on long follow-ups of samples of offenders. While both short and long follow-ups are relevant to treatment provision, short follow-ups would be more relevant to the targeting of time-limited restrictive and constructive measures such as NOMS's supervision work, and can be regularly calibrated on more recent samples to ensure that the predictor stays up-to-date.

The issue of whether actuarial tools should focus on the most harmful offences is dependent on whether specialisation exists within sexual offending. Certainly, many sex offenders are specialists in the sense that, by comparison with other offenders, they have a far higher rate of sexual reoffending but lower rates of other types of reoffending (Ministry of Justice, 2010a: see Table A8). It is also possible that they may specialise in particular types of sexual offending (Soothill, Fitzpatrick and Francis, 2009). A meta-analysis of 2,630 online sex offenders, from nine studies, found that this group had 2.0% contact sexual and 3.4% rate of reoffences involving online indecent images of children, in follow-ups ranging from 1.5 to 6 years. It was concluded that it is likely that a distinct sub-group of online-only offenders exists (Seto, Hanson and Babchishin, 2010). There is certainly sufficient evidence in favour of specialisation within sexual offending to indicate that directed investigation of contact sexual reoffending is a worthwhile goal.[165]

The aims of the analysis set out in this chapter were therefore to:

1. Explore a recent dataset of NOMS offenders for evidence of specialisation in sexual reoffending involving direct victim contact ('contact sexual reoffending'), examining whether important risk factors can be differentially predictive for various types of sexual reoffending.
2. Construct statistical models to predict contact sexual reoffending.
3. Use these models to create a new predictor, the OASys Sexual reoffending Predictor (OSP), a tool which can be used easily by prison/probation and police practitioners. As the police have limited access to information on dynamic risk, and NOMS may sometimes

[164] The argument against including such offenders in the predictors is that the factors which explain offending by children are very heterogeneous and may be different to those which explain offending by adults. However, such differences in psychological factors are unlikely to neutralise the value of wholly or mostly static actuarial predictors to the extent that it would be better to not make a prediction. While risk assessments exist for juvenile sex offenders (e.g. Griffin, Beech, Print, Bradshaw and Quayle, 2008), no tools have been developed for adults whose sexual offending was restricted to childhood.

[165] Other actuarial risk assessments for sexual recidivism have been developed. They are summarised, and compared through meta-analysis, by Hanson and Morton-Bourgon (2009). None differentiate between types of sexual offence in their predicted outcome.

need to rapidly screen offenders, the role of dynamic risk factors in the predictor needs to be carefully considered.

4. Validate this new predictor in a robust manner, and compare it with RM2000/s, to determine whether there is likely to be genuine value in its implementation.

10.2 Approach

Important concepts

This study defines four groups of sexual offences.[166] Two groups cover contact offences, and two almost exclusively cover non-contact offences; they are thus sometimes aggregated as 'contact sexual' or 'non-contact sexual' respectively:

- The two contact offence groups cover offences against children ('contact child'), and those whose victims are known to be adults or are of unknown age ('contact adult'). Note that statutorily-defined grooming offences are included in the contact child group, as the motivation of this offence is to make sexual contact with a child.
- Offences involving the making, distribution, showing and advertisement of indecent images of children form one non-contact group ('indecent images').[167] The remaining group ('paraphilia') includes offences resulting from a range of sexual interests that are usually most easily gratified through offending, the most common of which to result in criminal conviction tends to be exhibitionism. Some of the other offences in the group are related to voyeurism and zoophilia.[168]

A distinction is also made between ***statutory*** and ***element/motivation*** sexual offences. A statutory sexual offence is one which is defined as sexual on the basis of the legal charge for which the offender has been cautioned or convicted. An element/motivation offence is one which is not a statutory sexual offence but has been identified as sexual on the basis of professional judgement that the offending behaviour included a sexual element or was sexually motivated. For example, offences such as rape, sexual assault, incest or indecent exposure are all statutory, whereas convictions for theft or actual bodily harm could only be classified as sexual on the basis of clinical observation of an

[166] NOMS now maintains documentation on which statutory offences are included in each offence group. While some offences are not straightforward to classify, the general approach taken is that victims known to be aged under 16 are considered to be children and other victims are not.

[167] These 'indecent images' offences are often referred to as internet offences. Technically, the Internet is not necessarily a component of these offences, which date back to section 1 of the Protection of Children Act 1978 and section 160 of the Criminal Justice Act 1988 – the defined age of the child rose from 16 to 18 under the Sexual Offences Act 2003. Clinical experience shows that virtually all offences charged under the relevant statutes do involve the Internet, so to describe them as such is not problematic. However, many other types of sexual offence with child victims, such as meeting a child following sexual grooming, causing a child to witness a sexual act or inciting a child to engage in sexual activity, can sometimes or occasionally involve the internet but on other occasions occur entirely offline. These have been coded as contact child offences. It is not therefore feasible to construct a coherent and comprehensive group of internet-related offences using statutory offence codes, which are the basis of OSP's offence classification. It appears more productive, for predictive purposes, to describe the indecent images offences using their statutory label, and set aside the question of whether these or any other offences were committed wholly or partially online.

[168] Zoophilia is a paraphilia involving cross-species sexual activity between human and non-human animals or a fixation on such practice. 'Zoophilia' and 'bestiality' have been distinguished on the basis that the former describes the desire to form sexual relationships with animals, whilst the latter describes the sex acts alone.

element or motivation. Where an offender's only known sexual offending history was a current element/motivation offence, they are referred to as an 'element/motivation only' offender/case.

This study uses ***survival analysis***, with the follow-up starting on the day of an offender's conviction leading to a community sentence or upon discharge from custody for their index offence. It then continues until the offender either:

- reaches the cutoff date without reoffending;
- is imprisoned for any offence;
- is recalled to custody; or
- commits the offence type being studied.

It can then be seen whether reoffending occurs in each month at risk. The ***hazard*** of reoffending is the probability that reoffending for the offence type of interest occurs during some short time period (in this study, a day), given that the offender has not already reoffended in this way.

The ***cutoff date*** for this study was 3 June 2010. Data were drawn from the Police National Computer (PNC) on 3 December 2010; this allowed dates of reoffending and at-risk periods until six months previously, as an offence committed after this date would too often have not yet resulted in a PNC-recorded conviction. At-risk periods were therefore 'censored' (cut off) at this point (if imprisonment or recall to custody did not lead to earlier censoring). Each offender had their own at-risk period: the number of months from sentence/discharge until 3 June 2010 or their imprisonment (for recall or any offence). In this study, at-risk periods ranged from one month to over six years. Imprisonment was counted on the basis of the date of sentence, but reoffending on the basis of the date of offence.[169] Imprisonment could either be for a new non-sexual offence or for a pseudoreconviction, an offence of any type committed before the start of the follow-up period but brought to justice after. For recalls, the date upon which the offender actually returned to custody was used.

Two forms of survival analytic data models were used to estimate the relationship between risk factors and reoffending over time, and consequently develop OSP. Cox proportionate hazards regression ('***Cox regression***') is a flexible method for determining whether risk factors increase or decrease the hazard of reoffending on any given day, but it does not produce predicted probabilities of reoffending. Accelerated failure time models using the Weibull distribution ('***Weibull regression***') do produce predicted probabilities; in these models, risk factors increase or decrease the expected number of days to reoffending.

[169] The use of sentence date meant that imprisoned offenders were only removed from the follow-up at the point at which we were sure they were no longer at risk of committing further offences, as it is not possible to account for periods on remand.

Validation of new prediction models is vital (Steyerberg, 2009). The ability of the risk predictor to distinguish reoffenders from non-reoffenders on the data used to develop the model is necessary but not sufficient. It is also necessary to determine whether the predictor will work on new offenders, as a statistical model which is ***overfitted*** (inappropriately complex) can include considerable ***optimism*** (loss of predictive power when applied to new offenders). This chapter therefore presents results covering both offenders in the initial sample and 'new' offenders. The main statistic used to measure predictive validity is the ***Concordance Index*** ('C Index'; Harrell, Lee and Mark, 1996). Meta-analyses of predictors of sexual reoffending (e.g. Hanson and Morton-Bourgon, 2009) produce mean Area Under Curve (AUC; the equivalent of the C Index for fixed follow-up periods) results of about 0.7 for the prediction of all sexual reoffending, but prediction of contact-only reoffending has been seldom attempted in formal statistical models.

Bootstrapping is a statistical technique recommended to estimate the optimism of predictive models (Harrell *et al.*, 1996). In this study, it is used to reduce the C Index of OSP to allow fair comparison with RM2000/s. In bootstrapping, many new samples of offenders, of equal size to the original sample, are created by selecting from the original sample with replacement;[170] these are used as described in the Procedure section below. The underlying assumption is that the original sample is a good representation of the range of offenders to be found in the population of all future sex offenders, and that sampling with replacement is therefore a reasonably close equivalent to gathering new samples of offenders from that population.[171]

Risk Matrix 2000

Risk Matrix 2000 (RM2000) is an empirically-derived actuarial risk assessment tool that uses historical information about offenders to divide them into categories that should differ substantially in their rates of reconviction for sexual or other violent offences. It was developed for use in the United Kingdom with males aged 18 and over who have been convicted of a sexual offence, at least one of which must have been committed when aged 16 or over. RM2000 has three scales:

- RM2000/s is a prediction scale for proven sexual offending;
- RM2000/v is a prediction scale for proven non-sexual, violent offending; and
- RM2000/c is a combination of the first two scales and predicts proven sexual or non-sexual violent offending.

[170] For example, suppose that we have an original sample of just four offenders, named A, B, C and D. One bootstrap sample might contain two copies of offenders A and D, and none of B and C. Another might contain one copy each of A and B, two copies of C and none of D. For larger sets of offenders, the number of different possible bootstrap samples is effectively infinite, and each possible sample would produce a different statistical model. (The magnitude of the differences between these models is important in estimating optimism.)

[171] Bootstrapping "may seem invalid [but] both theoretical and evidence supports this process" (Steyerberg, 2009, p.95). It is frequently used in medical statistics but is novel in correctional criminology and forensic psychology.

As NOMS has alternate, validated methods for predicting non-sexual violent reoffending,[172] this study does not refer further to RM2000/v or RM2000/c.

RM2000/s 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 were followed for two years. It was constructed using existing research knowledge to identify which individual factors should be incorporated into the tool and the weight that should be assigned to them. Two validation samples were available for RM2000/s: offenders who had engaged in treatment in custody (*N* = 647, follow-up of 2+ years), and offenders discharged from prison in 1979 (*N* = 427, follow-up of 19 years). These samples yielded AUC statistics between .75 and .77 (Thornton *et al.*, 2003), although they have been criticised for a possible lack of representativeness (Grubin, 2008). Hanson and Morton-Bourgon's (2009) international meta-analysis found the RM2000/s scale to have moderate predictive accuracy for sexual reconviction (*d* = 0.67, which equates to an AUC of .68; Rice and Harris, 2005).[173] Barnett *et al.* (2010) found a similar result (AUC of .68) in a two-year proven sexual reoffending follow-up (*N* = 4,946). Wakeling, Howard and Barnett (2011) considered validity among a subsample of known indecent images (described as internet) offenders, finding that those rated very high risk had an elevated risk of proven sexual reoffending (AUC = .67, *N* = 994). The latter two studies combined OASys and treatment cases using earlier versions of the databases used in this study.

RM2000/s comprises seven items divided into two scoring steps. Step One includes three items:

- Age of the offender on release.
- Number of sentencing occasions for a sexual offence.
- Number of sentencing occasions for any criminal offence.

The scores assigned to each of these items are summed and translated into one of four preliminary risk categories: low, medium, high or very high. Step Two has a further four risk-raising items:

- Any of the victims of sexual offending have been male.
- Any of the victims of sexual offending (excluding internet offences) have been strangers.
- The offender has never had a continuous live-in relationship for over two years.
- Any of the sexual offences have been non-contact (excluding internet offences for those whose only sexual offences relate to the internet).

[172] The Offender Group Reconviction Scale (OGRS) score and RM2000/v have been shown through meta-analysis to have similar levels of predictive validity for non-sexual violent recidivism (Yang, Wong and Coid, 2010). The OASys Violence Predictor has been shown in a large validation sample of NOMS offenders to have superior predictive validity to both of these tools (Howard, 2009). Version 2 of OGRS was used in Yang *et al.*'s meta analysis, but the superior version 3 (OGRS3; Howard, Francis, Soothill and Humphreys, 2009) is now used in NOMS. RM2000/v was originally developed for use with known sex offenders, but has been validated for use with other offenders.

[173] Finding on the predictive accuracy of other actuarial and clinical risk assessment tools are summarized in Craig, Beech and Harkins (2009).

These items are scored on a dichotomous scale as either present or not. If two or three of these items are present the initial risk category is raised one level (e.g. from low risk to medium). If all four of these aggravating factors are present the initial risk category is raised by two risk levels (e.g. from low to high).

Sample

Lists of offenders assessed using OASys by 31 March 2008 were submitted to the Ministry of Justice's (MoJ's) PNC research database.[174] Cases were matched while applying rigorous consistency checks on OASys and the PNC's records of offenders' index offence conviction or sentence date, date of birth and gender. The matched offenders included 14,804 who were retained in the OASys male sex offender sample. These either had:

(i) incurred at least one sanction (caution or conviction), before their follow-up start date, for statutory sexual offending; and/or

(ii) an index offence marked as having a sexual element (OASys question 2.2F) or motivation (OASys question 2.9).

In counting sexual sanctions, whether overall or for the four sub-groups or two overarching groups defined above, the index sanction could be counted as sexual if it was not statutorily sexual but had sexual element or motivation indicated. 960 of the 14,804 offenders met this criterion and were element/motivation only. The OASys sample could not include offenders whose only sexual offending was both (i) prior to the index offence, and (ii) of an 'element/motivation' rather than statutory nature. While this appears to be a shortcoming of OASys compared with RM2000/s, the authors' understanding from clinical experience and user consultation is that such offenders are seldom identified as sexual offenders as detailed information regarding previous offences is often difficult to obtain. Moreover, this element of RM2000/s is time-consuming to score. The omission of previous element/motivation offences, for those offenders who do qualify for the sample, is an acknowledged weakness in terms of pure prediction, but should aid relatively quick and reliable scoring.

To assist with the simulation of RM/2000s scores (see below), access was gained to a further treatment database held within NOMS. This database holds administrative and risk assessment information on men who have undertaken a sex offender treatment programme in custody in England and Wales since 2001. The treatment sample included 1,047 cases (not necessarily discharged) meeting the following criteria:

- completed treatment by early 2008;
- not on the OASys database;
- matched with the PNC database; and
- had complete RM2000/s assessments.

[174] A number of filters on OASys data quality and completeness and checks for duplication were completed.

The use of this sample in the study was relatively limited (see Procedure), so it is not described in detail.

Female offenders were not represented in the treatment database, and were not retained in the OASys sample. Existing evidence suggests that the aetiology of their offending differs from that of males (e.g. Nathan and Ward, 2001). Scrutiny of the 223 female OASys cases showed that around half involved element/motivation only offenders, with child neglect among the most frequent such index offences. The rate of proven sexual reoffending was extremely low.

Of the 14,804 cases remaining in the OASys sample:

- 13,295 (90%) could be followed up for 24 months;
- 9,483 (64%) for 36 months;
- 5,795 (39%) for 48 months; and
- 2,753 (19%) for 60 months.

The mean follow-up length was 37.3 months, with a standard deviation of 20.3 months. The mean age of the offenders was 38.1, with a standard deviation of 13.6. Approximately two-fifths (39%) were on licence from a custodial sentence. Principal index offences were statutory sexual for approximately half (48%) of the cases, element/motivation for 9% and non-sexual for 42%.

Procedure

Previous sanctions, proven reoffending and recalls

Records of offenders' known criminal history and proven reoffending were extracted from the MoJ's PNC research database on 3 December 2010. Proven reoffending comprised offences committed between each offender's date of community sentence or discharge from custody and 3 June 2010, leading to caution or conviction in this period or the six-month 'buffer period' (set aside to ensure that most court cases and PNC data entry could be completed). Recall dates were identified by matching offenders with an extract from NOMS's Public Protection Unit Database. Statutory sexual offences were identified, and subcategorised into the contact adult, contact child, paraphilia and indecent images groups, on the basis of established offence descriptions (Ministry of Justice, 2010b). Counts of previous sanctions added past statutory sanctions to index statutory sanctions and index element/motivation sanctions. Where a previous sanction involved multiple offences from more than one of the four sexual offence sub-groups, it was classified using the primary offence (this is coded on the PNC data extract, on the basis of sentence severity).

The survival analysis methods used for statistical modelling meant that data for each offender could be used over their entire follow-up period, rather than using a fixed-length follow-up which would use data less efficiently. Using the censoring rules described above to determine offenders' follow-up periods, the OASys sample included 520 sexual reoffenders, including 269 contact reoffenders. The treatment sample included 60 and 31 such reoffenders respectively. Given that validation of prediction models should involve at least 100 cases with the event of interest (i.e. reoffenders; Steyerberg, 2009), the treatment sample was not used for validating the model developed on the OSP construction sample. Given its lack of dynamic risk factors, the treatment sample could also not be used as part of the OSP construction sample. The treatment sample was instead used as a source of information on the joint distribution of RM2000/s risk factors in order to simulate RM2000/s scores in the OASys sample, as described below.

Selection of risk factors

Static risk factors were coded from PNC data; all of these could be coded by practitioners or administrative staff on the basis of summary printouts of individual offenders' demographics and criminal histories. Such printouts are routinely available from the operational PNC. To assess the additional value to be gained from incorporating dynamic risk factors, relevant items were selected from OASys's offending-related factors component. The proposals on meaningful risk factors made by Mann, Hanson and Thornton (2010) were translated into OASys items, then reduced to those considered most theoretically and empirically promising through consultation with NOMS's Sex Offender Treatment Programmes (SOTP) team. The number of static and dynamic risk factors coded was deliberately restricted, as research into predictive model construction and validation suggests that there should be at least ten events for every candidate predictive risk factor (Harrell *et al.*, 1996).

Data quality on most OASys items was good (Morton, 2009b), but the victim's relationship with the offender was not always noted. For the most frequent current contact sexual offences in the OASys sample, such data were missing in between 27% (rape, gross indecency with children) and 37% (sexual assault on male) of cases. Victim age data was necessary to identify child victims of element/motivation offences, and was also sometimes missing. (For statutory sexual offences, the PNC offence code identifies whether there was a child victim.)

Specialisation within sexual offending

Combinations of histories of offending in the four sexual offence sub-groups were calculated. While there were several hundred different combinations, the majority of offenders fell into a small number of combinations. Proven reoffending rates for each of the four sexual offence sub-groups were calculated for each of the most frequent combinations. In addition, Cox regression models were developed to predict each of the four types of reoffending separately.

Prediction of contact sexual reoffending

On the basis of the specialisation results, a decision was taken that OSP should predict overall contact sexual reoffending: it was evident that risk factors for contact offences were different to those for non-contact offences, while the numbers of reoffenders were too low for separate, robust models of contact adult and contact child reoffending to be developed.

The removal of non-contact offences from the outcome lowered the number of events (reoffenders) available for model construction. The 10:1 events/predictors ratio was not achieved, as there were several possible predictive static items across the domains of age, desistance from sexual offending, sexual offending history and general offending history, while the SOTP team consultation similarly produced a moderately long list of potentially valuable dynamic items. Instead, the 269 reoffenders compare to 33 candidate predictors, for a ratio of about 8:1.[175] The reduced number of events made it impractical to split the OASys sample into construction and validation halves, and the bootstrap procedure described below was instead used to indicate whether the large number of candidate predictors had resulted in a sufficiently stable model.

The predictive models were produced using Cox regression, with risk factors being selected through the backwards stepwise method (*p*=.15).[176] It was considered necessary to reduce the number of risk factors in the model, despite the statistical advantages of fitting full models (Steyerberg, 2009), as models with many risk factors would be difficult to apply in some situations (i.e. without computer support) and to interpret.[177] The face validity of the models was checked – to see, for example, if any supposed risk factors were scored as protective – and the fit of the selected models was checked using Weibull regression.[178]

The selected regression model outputs were transformed into point-based scoring systems, by a manual process involving identification of the highest common factors among possible scores on each selected risk factor's regression coefficients, and an iterative check on the risk categorisation properties. Weibull regression models were fitted to the scores, producing an equation allowing predicted probabilities of contact sexual reoffending to be estimated for any combination of score and follow-up period. Such probabilities were produced for each offender on the assumption that their follow-up period would last two years, enabling predicted probabilities to be produced in a familiar and easily understood format. The two-year contact sexual reoffending probabilities associated with each score could therefore be displayed, and the scores divided into low, medium, high and very high risk categories on the basis of these probabilities.

---

175 Three predictors were gained when variants of age at start of follow-up and number of non-sexual sanctions were produced, and age at most recent sexual sanction was split into "under 16" and "16-17" variables. In addition, the "years since last sexual offence" item shown in the Results was selected, through direct comparison of correlations with contact sexual reoffending, over an alternate version which would have counted the number of years between last sexual offence and current conviction and therefore not counted time in custody. This alternate version is not counted as a candidate predictor, as it was not used in any multivariate models.

176 This *p* value is higher than those cited in most other studies. Steyerberg (2009) suggests that the use of *p* values of .05 has been inappropriately transplanted from hypothesis testing to model selection. The *p* value of .15 is equivalent to the favoured AIC criterion of assessing model goodness of fit; this criterion rewards improvements in model fit but penalises loss of parsimony (increases in the number of items included).

177 A static/dynamic model was created using the risk factors selected in the static model, and adding dynamic risk factors which incrementally improved the model's predictive performance. It was recognised that using the same set of static risk factors would provide practical benefits including less complex user communication, briefer and easier user training, and reduced scoring effort in situations where offenders were 'upgraded' from the static to static/dynamic models.

178 All data analysis was conducted in SAS software version 9.2. SAS 9.2 allows automated backwards stepwise model selection when fitting Cox regression models in PROC PHREG. It does not allow automated model selection when fitting Weibull models in PROC LIFEREG. Given that there were 2^{33} possible combinations of risk factors, automated model selection was a necessity.

Simulation of Risk Matrix 2000/s scores

RM2000/s scores were simulated as closely as possible from the available data. As in Barnett *et al.* (2010) and Wakeling, Howard and Barnett (2011), it was possible to score Step One and the Step Two male victim and ever non-contact items, with the minor caveat over previous element/motivation offences explained above. Four final risk categorisations were then determined, based on the following simulation rule sets for the Step Two relationship and stranger victim items.

(i) Do not score relationship and stranger victim, and increase risk categorisation between Steps One and Two using the same rules as in Barnett *et al.* (2010) and Wakeling, Howard and Barnett (2011). The modified scoring rule involves raising the initial risk category by one level if the offender had one or both of the 'male victim' or 'ever non-contact' (with the above internet-only modification) aggravating risk factors.

(ii) Create OASys indicators of whether the offender is currently living with a partner ('current partner'; OASys question 3.2), and whether their current offence was sexual with a stranger victim ('current stranger victim'; OASys question 2.4).[179] Use these items in place of the missing Step Two items, and increase risk categorisation according to the standard RM2000/s procedure.

(iii) Estimate the incremental predictive contribution of the two OASys-based indicators described in rule set (ii), using a Cox regression model of contact sexual reoffending which also includes the Step One score. From the treatment sample, model the relationship between the two missing items and Step One score. Use the resulting probability distributions to simulate scores on the two missing items separately for reoffenders and non-reoffenders, stratifying on the Step One score, to reflect the estimated hazard ratios for the equivalent OASys-based items from the Cox model. (Apply the current partner item to relationship, and the current stranger item to stranger victim.)

(iv) As rule set (iii), but use the upper confidence limits of the Cox model's hazard ratios rather than the central estimate of the hazard ratios.

Rule set (i) is likely to underestimate the predictive validity of RM2000/s for contact sexual reoffending, as the missing Step Two items should be more predictive than the non-missing Step Two items. Rule set (ii) assumes that the OASys indicators are about as predictive as the missing Step Two items, in the sense that the hazard ratios between those with and without the risk factors are similar. This is reasonable, but may cause bias because the OASys indicators are less frequent than the missing items, as the former measure current or index offence statuses while the latter measure similar lifetime statuses. Rule set (iii) is also founded upon the assumption that the OASys indicators are equally predictive in terms of their hazard ratio, but it additionally uses information from the treatment sample to allow realistic simulation of the frequency and distribution of the missing RM2000/s items more

[179] Internet offences were excluded as in the RM2000 scoring guide (Thornton, 2007). RM2000 defines stranger victims according to whether the victim and offender mutually knew each other 24 hours before the offence. "Knowing minimally involves having physically met, had a conversation with, and being able to recognise the other person". OSP user guidance could clarify the application of these criteria to relationships based on online video conversation.

accurately. Rule set (iv) varies from set (iii) by assuming that the RM2000/s indicators are considerably more predictive than their OASys equivalents.

This range of simulation rules was used to test whether the missing RM2000/s items could be ignored without any loss of predictive validity (comparing the results for sets (i) and (iii)), and whether the OASys indicators could substitute for the missing items in a simple way without loss of performance (comparing results for sets (ii) and (iii)). If these tests show that set (i) and (ii)'s simple scoring rules result in underestimates of RM2000/s's predictive validity, sets (iii) and (iv) can be used to compare the predictive validity of OSP with RM2000/s. Set (iii) is considered by the authors to be the fairer of the two for this comparison as, if anything, the presence of risk on the OASys indicators should be a better discriminator between reoffenders and non-reoffenders than the presence of risk on the missing items, given that the OASys indicators refer to current status rather than a status which may be long past.[180] Set (iv) is included in order to offer stronger evidence that OSP improves predictive validity. That is, its scoring rules reflect an opposing view in which the superior definition of the missing items ought to grant them far greater predictive value than the OASys indicators.

Model validation

A set of 100 bootstrap samples,[181] each containing 14,804 cases (equal in number to the original sample), was produced. New prediction models were fitted on these samples using the same p value and face validity procedures as were used for the original OSP model, and cases were placed into risk categories of equal sizes to those derived from the original model.

Statistical modelling was undertaken on these bootstrap samples with the goal of making optimism estimates for OSP, to allow fair comparison of RM2000/s and OSP. The C Indices of the four RM2000/s variants were estimated on the original sample. The C Index of each new model was estimated both on its own bootstrap construction samples and on the original sample. In standard bootstrapping methods, the mean decrease in predictive validity between these bootstrap construction samples and the original sample forms the optimism estimate. In this study, a strong correlation was observed between bootstrap construction C Index and optimism, meaning that the amount of loss of predictive power due to optimism was much greater for bootstrap models with high construction C Indices. These models typically included more of the candidate risk factors, and evidently were less robust than the other, simpler bootstrap models with lower construction C Indices. To account for this potential bias, a linear regression model was fitted to allow OSP's optimism estimate to be conditional upon the C Index found for the OSP construction sample. This extra step ensured that the bootstrap procedure fulfilled its stated purpose: to estimate the amount of optimism typically found when models similar to the original model are fitted.

[180] Non-violent and non-sexual violent reoffending results demonstrate that other OASys items (i.e. those scored in OGP and OVP) are genuinely dynamic and that score changes on these items are incrementally predictive of reoffending in periods following the score changes – see Chapter 4. This suggests that current status is preferable, in predictive terms, to past status.

[181] The number of samples used was capped at 100 due to the intense computational demands of the modelling process.

Sensitivity and specificity comparisons were produced for the grouped OSP predictor and rule set (iii)'s version of RM2000/s. These measures show how well the predictors separate reoffenders from non-reoffenders, as applied in practical settings where risk predictor categories will commonly inform decisions about which offenders will be the target of enhanced treatment, supervision and/or surveillance.[182]

10.3 Results

Specialisation within sexual offending

Table 10.1 compares offence-specific reoffence rates, controlling for offenders' previous sanction counts for each offence type in turn. While overall reoffending rates were relatively similar across the four offence types, evidence of specialisation is clear – for all except contact child offending, those with the most extensive history of an offence type had at least 20 times the offence-specific reoffence rate of those with none. For the two contact offence groups, those with one sanction for the offence had about twice the offence-specific reoffending rate of those with none, whereas specialisation in the two non-contact groups was far stronger.

- The 36% of the sample with some history of contact adult offending provided 66% of contact adult reoffenders.
- The 49% of the sample with some history of contact child offending provided 67% of contact child reoffenders.
- The 15% of the sample with some history of paraphilia offending provided 72% of paraphilia reoffenders.
- The 10% of the sample with some history of indecent images offending provided 53% of the indecent images reoffenders.

Only a small proportion of offenders had been sanctioned for any of the offence types on more than three occasions. Paraphilia did though show a long-tailed distribution, in that a few offenders had very many sanctions, and it can be seen that the ratio of those with two to three or more offences was very different than for the other offence types. This is indicative of extreme levels of repetition of this offence, and therefore extremely high reoffending rates, among a relatively small sub-group of offenders.

[182] These results are slightly affected by differences in time at risk between lower and higher risk offenders, but all predictors were affected in the same way.

Table 10.1: History of each sexual offence and proven reoffending rates for four types of sexual offence

Previous sanction Count	Number of offenders	% proven reoffending by each offence type: Contact adult	Contact child	Paraphilia	Indecent images
All offenders	14,804	1.1%	0.7%	1.1%	0.8%
Contact offences with adult victims					
0	9,413	0.6%	0.9%	1.2%	1.2%
1	4,797	1.6%	0.5%	0.8%	0.1%
2	462	3.9%	0.6%	2.6%	1.1%
3 or more	132	12.1%	0.8%	0.8%	0.8%
Contact offences with child victims					
0	7,618	1.5%	0.5%	1.6%	0.9%
1	6,190	0.8%	0.9%	0.6%	0.5%
2	704	1.1%	1.6%	0.4%	1.8%
3 or more	292	1.0%	2.4%	1.7%	3.8%
Paraphilia offences					
0	12,512	1.2%	0.8%	0.4%	0.8%
1	1,855	1.0%	0.3%	3.1%	0.8%
2	258	1.6%	1.9%	10.5%	1.2%
3 or more	179	2.2%	1.7%	18.4%	1.1%
Indecent image offences					
0	13,334	1.3%	0.8%	1.1%	0.4%
1	1,392	0.2%	0.7%	0.9%	4.2%
2	72	0.0%	0.0%	0.0%	9.7%
3 or more	6	0.0%	0.0%	0.0%	16.7%

Table 10.2 shows the reoffending rates associated with particular combinations of sexual offence histories. All combinations of one or two previous sanctions are shown, together with other combinations involving at least 50 offenders, with the final row ensuring that all 14,804 offenders are represented. Comparison of the four rows covering those with a single offence type shows the extent of specialisation: in each row, the previous sexual offence was the most frequent reoffence, and for three of the four types it was more frequent than the other three combined. Among offenders with two previous sanctions, those with exclusively non-contact histories showed extreme specialisation, with high levels of non-contact reoffending and little contact reoffending. In RM2000/s, ever having committed non-contact sexual offences is an aggravating factor for all sexual reoffending: this appears to be true in the sense that the overall level of sexual reoffending is high, but it is restricted to non-contact offences. Those with contact adult histories showed considerable specialisation, while those with contact child histories sometimes ‘crossed over’ to indecent image reoffending. Strong specialisation patterns were present among those with three previous contact adult, contact child or paraphilia sanctions.

The numbers with each combination of sanctions provide further evidence of specialisation. If no specialisation occurred, then the number of offenders with one contact child and one paraphilia sanctions would vastly exceed the number with two paraphilia sanctions, as contact child sanctions are more frequent overall, yet in fact the latter group is larger. Similarly, among contact offenders with two or three sanctions, there are fewer offenders with histories of both offences than statistical independence would allow.

Considering tendencies to specialise in contact sexual reoffending, Table 10.2's results can be summarised as follows:

- The 75% of the sample with only contact offending history comprised 81% of the contact reoffenders but just 28% of the reoffenders with only non-contact reoffences.
- The 20% of the sample with only non-contact offending history comprised 12% of the contact reoffenders and 51% of the non-contact-only reoffenders.
- The 5% of the sample with histories of both contact and non-contact offending comprised 7% of the contact reoffenders and 22% of the non-contact-only reoffenders.

Table 10.2: Combinations of sexual offending histories and proven reoffending rates for four types of sexual offence

Previous sanctions for sexual offences				Number of offenders	% proven reoffending by offence type			
Contact adult	**Contact child**	**Paraphilia**	**Indecent image**		**Contact adult**	**Contact child**	**Paraphilia**	**Indecent image**
Offenders with one previous sexual offence sanction								
1				3,980	1.6%	0.4%	0.5%	0.1%
	1			5,410	0.6%	0.9%	0.2%	0.4%
		1		1,368	0.9%	0.2%	2.8%	0.4%
			1	1,176	0.2%	0.7%	0.5%	3.7%
Offenders with two previous sexual offence sanctions								
2				334	4.2%	0.9%	1.2%	0.6%
1	1			407	2.0%	1.7%	0.5%	0.2%
1		1		176	1.7%	0.6%	3.4%	0.0%
1			1	19	0.0%	5.3%	0.0%	0.0%
	2			554	0.9%	1.4%	0.0%	2.2%
	1	1		133	0.0%	0.8%	3.8%	1.5%
	1		1	71	0.0%	0.0%	1.4%	4.2%
		2		179	1.1%	1.7%	10.6%	0.0%
		1	1	61	0.0%	1.6%	8.2%	9.8%
			2	54	0.0%	0.0%	0.0%	11.1%
Offenders with three or more previous sexual offence sanctions ***(n>50 only)***								
3				76	11.8%	1.3%	0.0%	0.0%
1	2			57	1.8%	1.8%	0.0%	0.0%
	3			158	0.6%	3.8%	1.3%	1.9%
		3		87	2.3%	1.1%	16.1%	0.0%
All other combinations				*504*	*3.2%*	*0.8%*	*6.3%*	*2.8%*

Table 10.3 displays the Cox regression models fitted for each type of sexual reoffending. Contact adult and paraphilia offences appeared to be more associated with general criminality than contact child and indecent image offences, as the first-time entrant item was strongly associated with the former but not the latter. The age patterns for the two contact groups both predict a simple reduction in risk as age falls. The quadratic age term for paraphilia predicts a very strong reduction in risk as age falls, but this is offset by the age at first sanction term. The age pattern for indecent image offenders indicates a peak in risk around age 35, falling off more sharply at the oldest ages. Risk may persist more over time for the two contact groups: the juvenile offending terms were less pronounced for these groups, while the 'years since last sexual offence' term indicated that desistance effects were stronger for non-contact but not contact offences. Having ever had a male sexual offence victim (based on PNC offence codes, which for sexual offences routinely separate offences with male victims from those with female victims) was predictive of non-contact but not contact reoffending. The stranger victim item, restricted as stated above to contact index offences, was associated with raised risk of contact adult reoffending. In RM2000/s, both of these items are aggravating factors for all sexual reoffending.

Table 10.3: Cox regression models of four types of sexual reoffending

	Model outcome (n/% of N=14804 reoffending rather than censored)			
Risk factor	**Contact adult** **170 (1.2%)**	**Contact child** **111 (0.8%)**	**Paraphilia** **164 (1.1%)**	**Indecent image** **125 (0.8%)**
General criminal history				
Sanctions for non-sexual offences	NS	NS	NS	NS
Square root of sanctions for non-sexual offences	NS	NS	NS	NS
First-time entrant?	-1.03***	NS	-0.98**	NS
Age and history of offending items				
Age at start of at-risk period	-0.032***	-0.031***	-0.084*	0.070*
Age at start of at-risk period, squared	NS	NS	0.00063*	-0.00098*
Age at first sanction	NS	NS	0.019*	NS
Last sexual sanction when aged <16	-1.37**	-1.13*	-2.39*	-1.63*
Last sexual sanction when aged 16-17	-0.60*	NS	-1.15*	NS
Years since last sexual sanction	-0.020*	NS	-0.044*	-0.075**
Sexual offending history				
Contact adult sanctions (max 3)	0.85***	NS	NS	-0.36*
Contact child sanctions (max 3)	NS	0.73***	-0.27*	0.51***
Paraphilia sanctions (max 3)	NS	0.38*	1.27***	NS
Internet sanctions (max 3)	-1.07*	NS	NS	1.59***
Ever committed sex offence with known male victim?	NS	NS	0.71**	0.62*
Current offence sexual with child family member victim?	NS	NS	NS	NS
Current offence contact sexual with a stranger victim?	0.51**	NS	NS	NS

Note. Models were fitted using backwards stepwise selection. ***: p<.001. **: p<.01. *: p<.15. NS: not significant at p=.15.

As in Tables 10.1 and 10.2, sexual offending history indicated strong specialisation, especially for non-contact offences. A history of paraphilia was somewhat predictive of contact child reoffending, though contact child history was somewhat negatively predictive of future paraphilia. Those with a history of contact child offending were more likely to commit new indecent image offences, but not vice versa. Indecent image offenders appeared very unlikely to commit contact adult reoffences. Of the sexual deviance items,[183] having a current sexual stranger victim was predictive of contact adult reoffending only, while ever having had a male sexual victim was predictive of both non-contact groups but neither contact group.

The strong evidence of specialisation in the differing types of sexual offence displayed in Tables 10.1, 10.2 and 10.3 confirmed that OSP should be developed as a predictor of contact sexual offences only. Those whose histories were exclusively of non-contact offending were shown to be likely to reoffend sexually, but were far more likely to reoffend in this non-sexual manner rather than crossover to contact reoffending. Other risk factors, such as those related to desistance, age at start of follow-up and male victims, also predicted differently for contact and non-contact offences. Including non-contact sexual reoffending in OSP's outcome would therefore be likely to compromise prediction of the most harmful, contact sexual reoffending.

Prediction of contact sexual reoffending

Table 10.4 presents the statistical model used to create OSP, in its Cox and Weibull regression formats. The additional parameters in the Weibull model are used to estimate the shape of the hazard curve (i.e. how the probability of reoffending changes over the course of the follow-up) and its scale (i.e. the overall level of reoffending), whereas Cox regression does not seek to model the hazard.

The selected model covered all expected domains of static risk, including eight of the 16 candidate predictive factors. In the order shown in Table 10.4, the candidate predictive factors are grouped in the domains of:

- general offending history;
- age;
- desistance from sexual offending;
- sexual offending history; and
- sexual deviance.

Within general offending history, the selected factor was whether the offender was a first-time entrant – that is, had no known previous offending history. The risk of reoffending was much lower among those who had desisted from sexual offending since they were juveniles and especially if the most recent sexual offence was committed aged under 16. The sexual offending history domain was complicated by the evidence for specialisation: contact adult sanctions were most predictive, while

[183] In this chapter, the "sexual deviance" items focus upon offence-related sexual interests.

contact child and paraphilia sanctions were moderately predictive and indecent image sanctions were unpredictive. Of the hypothesised factors around victim choice and deviance level, the stranger victim item – which perhaps fits least well into any of the established domains – was associated with raised risk. The selected risk factors all had comparable parameter estimates in the Weibull model (remembering that Weibull parameters have the opposite sign to Cox parameters). The shape parameter of 1.39 indicates a hazard which was roughly constant in early time periods then declined later.

Table 10.4: Cox and Weibull regression models of contact sexual reoffending

Risk factor	Cox regression		Weibull regression	
	Parameter	**SE**	**Parameter**	**SE**
Sanctions for non-sexual offences	NS		n/a	
Square root of sanctions for non-sexual offences	NS		n/a	
First-time entrant?	-0.64**	0.20	0.89**	0.28
Age at start of at-risk period	-0.035***	0.006	0.048***	0.008
Age at start of at-risk period, squared	NS		n/a	
Age at first sanction	NS		n/a	
Last sexual sanction when aged under 16?	-1.30***	0.33	1.78***	0.48
Last sexual sanction when aged 16-17?	-0.51*	0.24	0.69*	0.34
Years since last sexual sanction	NS		n/a	
Contact adult sanctions (max 3)	0.64***	0.08	-0.91***	0.13
Contact child sanctions (max 3)	0.38***	0.08	-0.53***	0.12
Paraphilia sanctions (max 3)	0.27**	0.10	-0.38**	0.14
Indecent image sanctions (max 3)	NS		n/a	
Ever committed sexual offence with male victim?	NS		n/a	
Current offence is sexual with a child family member victim?	NS		n/a	
Current offence is contact sexual with a stranger victim?	0.51***	0.15	-0.73***	0.22
Weibull intercept			*11.41*	*0.39*
Weibull shape parameter			*1.39*	*0.08*
Weibull scale parameter			*0.72*	*0.04*

Note. The Cox model was fitted by backward selection. The Weibull model automatically included the same risk factors as the Cox model. ***: p<.001. **: p<.01. *: p<.15. NS: not significant at p=.15. SE: standard error.

Table 10.5 displays the scoring chart devised by rounding the estimated regression coefficients from the statistical model. The observed age range, which included some offenders aged over 70 and a small number aged over 80, was restricted by noting that the 95th percentile of the age distribution lay at age 63. The most important factor, at a maximum of 8 points, was the contact adult sanction count, followed by age at start of follow-up, at 7 points. The other factors ranged between 2 and 5 points.

The scoring chart also includes the risk categories. These were determined on the basis of the two-year recidivism rates which could be estimated using the Weibull model parameters. These are displayed in Table 10.6 (with Figure 10.1 displaying a survival chart for the data used to construct OSP). Before the predictive model was created, it was decided that the category boundaries should be set at approximately half of the mean rate, the mean rate, and twice the mean rate. The mean predicted two-year rate for the sample was 1.44%. About one-quarter of offenders had a predicted rate below half this level and were in the low risk category, with a mean rate of 0.45% which was below one-third of the overall mean. Over one-third had a predicted rate above half the overall mean but below the mean and were therefore medium risk. Under one-third had a predicted rate less than double the mean and were high risk. The final eight percent had a mean predicted rate of 4.57%, over three times the overall mean. The low risk category contained one-quarter of the sample and eight percent of expected reoffenders, while the opposite was true of the very high risk category.

Table 10.5: OSP scoring chart

Risk factor	Score								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Effective assessment age	60+	54–59	48–53	42–47	36–41	30–35	24–29	18–23	
Age at last sexual offence	<16			16–17		18+			
First-time entrant?	Yes			No					
Contact adult sanctions	0			1		2			3+
Contact child sanctions	0		1	2	3+				
Paraphilia sanctions	0	1	2	3+					
Stranger victim of current contact sexual offence?	No		Yes						
Total score (maximum score 32)									
Risk category (Low = 0–11; Medium = 12–14; High = 15–17; Very High = 18–32)									

Table 10.6: Expected two-year contact sexual reoffending rates and risk categories

OSP score or category	Number (%) of offenders	Predicted probability at this score / category (95% CI for categories)	Expected % of reoffenders at this score / category
All offenders	***14,804 (100%)***	***1.44%***	***100%***
5 or less	85 (0.6%)	0.14%	0.1%
6	93 (0.6%)	0.18%	0.1%
7	488 (3.3%)	0.23%	0.5%
8	433 (2.9%)	0.30%	0.6%
9	567 (3.8%)	0.38%	1.0%
10	880 (5.9%)	0.48%	2.0%
11	1,331 (9.0%)	0.62%	3.8%
Low (0–11)	***3,877 (26.2%)***	***0.45% (0.28%, 0.71%)***	***8.1%***
12	1,796 (12.1%)	0.78%	6.6%
13	1,708 (11.5%)	1.00%	8.0%
14	1,911 (12.9%)	1.27%	11.4%
Medium (12–14)	***5,415 (36.6%)***	***1.03% (0.79%, 1.33%)***	***26.0%***
15	1,931 (13.0%)	1.62%	14.7%
16	1,395 (9.4%)	2.07%	13.5%
17	1,011 (6.8%)	2.63%	12.5%
High (15–17)	***4,337 (29.3%)***	***2.00% (1.62%, 2.46%)***	***40.7%***
18	571 (3.9%)	3.34%	9.0%
19	283 (1.9%)	4.25%	5.6%
20	182 (1.2%)	5.39%	4.6%
21	59 (0.4%)	6.82%	1.9%
22 or more	80 (0.5%)	10.92%	4.1%
Very high (18–32)	***1,175 (7.9%)***	***4.57% (3.52%, 5.92%)***	***25.2%***

While the total possible range of scores is wide, the gap between the low and very high categories was seven points (from 11 to 18). About nine-tenths of offenders scored between 9 and 18 points. In practice, extremely low scores will be rare due to the presence of both the age at start of follow-up and age at most recent sexual sanction items: few offenders will be old at the start of follow-up despite having committed all sexual offences before the age of 18, and those who did would score some points for their offence history as their known sexual offending would have predated the internet era and therefore be less likely to involve indecent images. Most of those whose sexual offending took place before the age of 16 (11% of the sample) were rated low or medium risk [184] Extremely high scores will also be rare because, as Tables 10.1 and 10.2 showed, very few offenders had multiple

[184] OSP was found to be valid for this sub-group, which had a 0.6% (10/1610) contact reoffending rate. Given their zero score on the “age at last sexual offence” risk factor, it is unsurprising that most were rated low or medium risk, but contact reoffending rates within each group were consistent with the rates of the entire sample.

offences across multiple sexual offence categories, and the frequent imprisonment of repeat sex offenders means that any who did will likely have been relatively old at the start of follow-up.

Tables 10.6 also sets out the expected two-year contact sexual reoffending rates for each risk category. Confidence intervals were calculated by assuming a binomial distribution, using the exact formula specified by Mood, Graybill and Boes (1974) rather than the standard large-sample approximation.[185] Importantly, the confidence intervals associated with adjacent categories did not overlap, indicating that the categories delineate groups of offenders with genuinely different reoffending rates.

Simulation of Risk Matrix 2000/S scores

As described in the Procedure section, Step One of RM2000/s, and the ever male victim and ever non-contact aggravating factors, were scored in broad accordance with the RM2000 scoring guide. The simulation of RM2000/s scores using rules sets (iii) and (iv) was completed by imputing values for the aggravating factors of never having had a stable live-in relationship and ever having had a stranger victim. To estimate the relationship between these factors and contact sexual reoffending, a logistic regression model estimated the effect on contact sexual reoffending of the OASys current live-in partner[186] and stranger victim items, controlling for the RM2000/s Step One score. Both items were significant predictors, with central hazard ratio estimates of 1.63 for the partner item and 1.88 for the stranger victim item.[187] In the treatment sample, the distribution of the equivalent Step Two items was found for each Step One score. For rules set (iii), these data and randomly generated probabilities were used to simulate the distributions of the missing items. For example, 15.7% of those scoring zero on Step One in the treatment sample had never had a stable live-in relationship. In the OASys sample, the contact reoffending outcome among those scoring zero on Step One was 25/3016. It was calculated that a 1.63 hazard ratio would result in this sub-group if the probability of never having had a stable live-in relationship was set at 0.232 among reoffenders and 0.156 among non-reoffenders. For rules set (iv), the same procedure was followed, but using the upper hazard ratio estimates of 2.23 (partner) and 2.44 (stranger). In this way, contact reoffenders were given higher mean scores than non-reoffenders on the partner and stranger items (and therefore higher final RM2000/s risk categories through rules set iii) and iv)), using evidence on the likely strength of these risk factors,

In fitting these scores, it was observed that the treatment sample's relationship item had an extreme age trend. While 89% of those aged 18–24 had never had a stable live-in relationship lasting at least two years, 45% of those aged 25–34 and 18% of those aged 35 and over had never had such a relationship. By focusing on current rather than lifetime relationship status, the OASys item reduces

[185] The exact formula correctly produces asymmetric confidence intervals. For example, the low risk group has a confidence interval of (0.27%, 0.65%), which is (-0.14%, 0.24%) compared with the central estimate of 0.41%.

[186] In OSP, this item is coded as current live-in nonvictim partner. In RM2000/s, no adjustment is made for whether the offender had victimised any partner with whom they had lived. Therefore, in simulating RM2000/s scores, the OASys item was adjusted to take no account of partner/victim status, in order to minimise differences with the RM2000/s definition.

[187] The Step One item was a significant predictor, with a hazard ratio of 1.58 per point.

this built-in age bias: in the three age groups, 89%, 79% and 77% respectively were not currently living with a non-victim partner. Therefore, as the hazard ratio was estimated incrementally with the Step One score, which includes age as a major risk factor, it is likely that use of the OASys item led to overestimation of the hazard ratio of the true RM2000 item. This strengthens the caveat, expressed in the Approach, that rule set (iii) may produce a generous estimate of RM2000/s's predictive validity, and rule set (iv) may produce a considerable overestimate.

Figure 10.2 displays a survival chart for the rule set (iii) version. Compared with Figure 10.1, the clear visible difference is the greater survival of the very high group.

Figure 10.1: Survival chart for contact sexual reoffending: categories of OSP

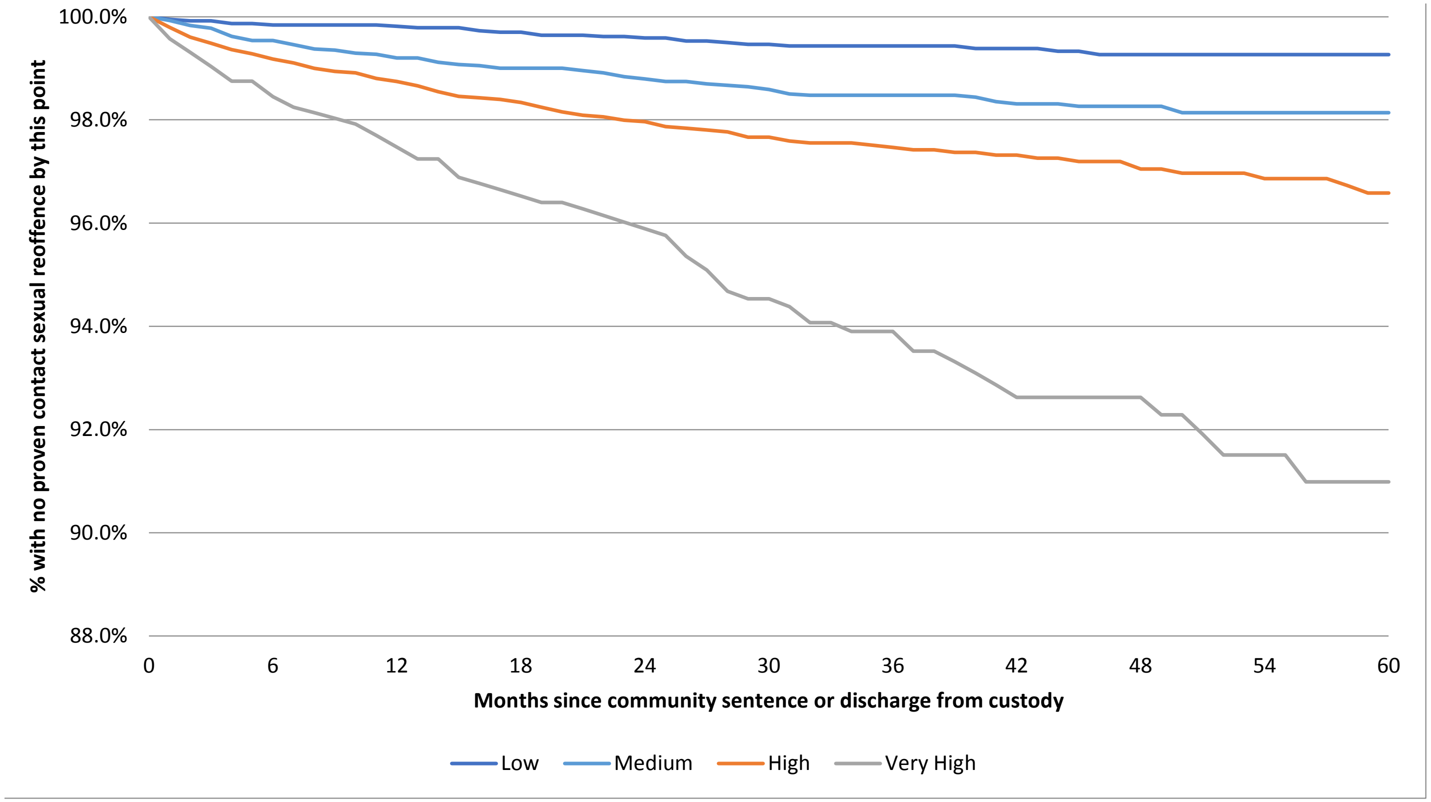
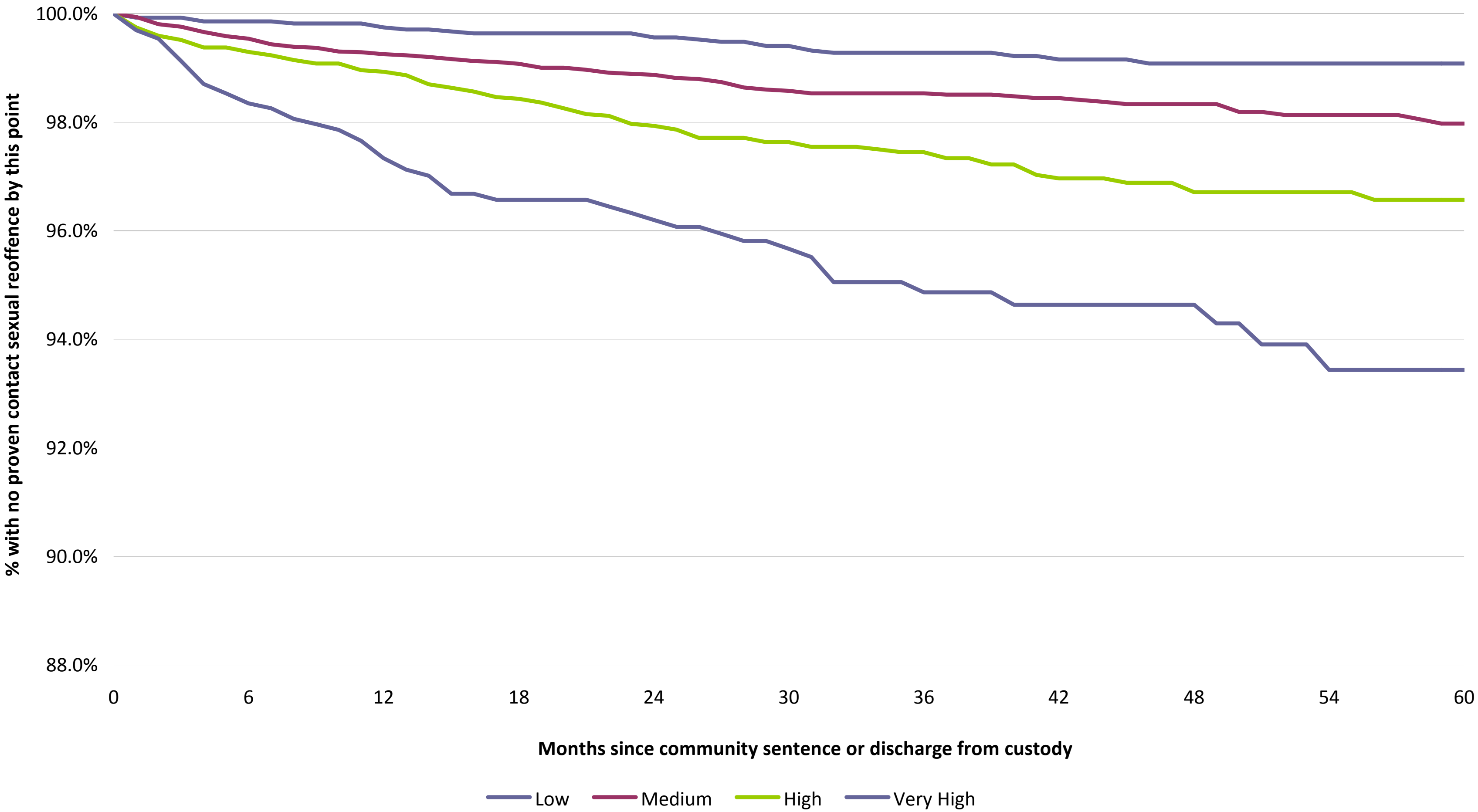


Figure 10.2: Survival chart for contact sexual reoffending: simulated Risk Matrix 2000/s categories



Model validation

Bootstrap sample analysis generated linear regression models, recognising that the optimism of a risk predictor is heavily dependent on its original performance, with strongly performing predictors (which typically are complex, involving many risk factors) shrinking most due to excess optimism. When these models were applied to the C Indices of OSP, they estimated that predictive performance should only decline by .005 when applied to new offenders. Shrinkage due to optimism was slightly greater for the ungrouped scores. In short, the OSP model was found to be a conservative model which could be expected to work almost as well on future offenders as they did on the current sample.

Table 10.7 displays Concordance Indices generated for OSP and RM2000/s, including estimates to adjust for OSP's optimism. The grouped version of OSP should be compared with RM2000/s; the ungrouped version is included to allow comparison with any ungrouped prediction score if necessary in the future, and to illustrate that predictive validity is always lost when moving from the raw score to the more practically useful grouped score. The optimism-adjusted results show that OSP was better able to discriminate reoffenders from non-reoffenders than RM2000/s. The size of the improvement in predictive validity depended considerably on the way RM2000/s scores were estimated, with OSP's advantage ranging from .07 to .006. Given the caveats expressed above, it is unlikely that true RM2000/s scores would have achieved more than the 0.654 achieved by rule set (iii); if this rule set is accepted, then OSP improved concordance by .022.

Comparing the grouped version of a combined static/dynamic OSP predictor to the grouped static only predictor, the size of the improvement in predictive validity was relatively small at .009.[188] Bearing in mind the additional resource and training required for scoring dynamic items and the complexities which would result when offenders had been assessed through two differing models, it was recognised that the static/dynamic model was not worth pursuing. It is probable that information on dynamic risk factors specific to sexual reoffending, rather than the general risk factors included in this study, is necessary in order to improve upon the predictive validity of a well-chosen set of static risk factors.

[188] The addition of dynamic factors to the static factors resulted in an extra four items entering the model. Two accommodation factors were added, so that offenders who lived with a non-victim partner were lowest risk and those of no fixed abode were highest risk. Current psychiatric problems and poor motivation also raised risk.

Table 10.7: Predictive validity of OSP and RM2000/s, with adjustments for model optimism

Predictor	Concordance Index	
	Before adjusting for optimism	**Optimism-adjusted, for fair comparison**
OSP, ungrouped	*0.696*	*0.694*
OSP, grouped	0.681	0.676
RM2000/s, rule set (i)	n/a	0.606
RM2000/s, rule set (ii)	n/a	0.635
RM2000/s, rule set (iii)	n/a	0.654
RM2000/s, rule set (iv)	n/a	0.673

Note. The RM2000/s scores were developed on a different sample, so their C Index has already been affected by elimination of the optimism effect.

Table 10.8 presents the practical implications of the new predictors' risk categorisation, by comparing the sensitivity and specificity of OSP and the rule set (iii) simulation of RM2000/s. These results slightly underestimate validity in an operational context, as they do not account for the shorter mean at-risk period of higher risk offenders (as noted in the Approach), and the sensitivity of OSP therefore appears lower than demonstrated in Table 10.6. Table 10.8 does however allow fair comparison between the two predictors.

For each predictor, setting the operational cutoff above the Low risk category means a sensitivity of about 90% – that is, about nine-tenths of reoffenders are above this cutoff – with a small lead for RM2000/s. However, the two predictors differ greatly in specificity, with RM2000/s placing just 19% of non-reoffenders in the low risk category compared with 27% for OSP. When the cutoff is set above the medium risk category (i.e. so low and medium risk offenders have lesser priority and high and very high risk offenders have greater priority), there is less difference in specificity, but sensitivity remains markedly different: 55% compared to 62%. If only very high risk offenders are prioritised, OSP still offers better sensitivity than RM2000/s. Relative risk ratios, set out in Table 10.9, also show that OSP produces greater differences between the proven contact sexual reoffending rates of low and medium risk offenders, and between those of high and very high risk offenders.

Table 10.8: Sensitivity and specificity of OSP and RM2000/s categories

Risk categories / score range below cutoff	Risk categories / score range above cutoff	% of all offenders above cutoff	Sensitivity	Specificity
OSP				
Low	Medium, high, very high	73.8	90.7	26.5
Low, medium	High, very high	37.2	62.1	63.2
Low, medium, high	Very high	7.9	24.5	92.4
RM2000/s [simulated, rule set iii)]				
Low	Medium, high, very high	81.1	91.5	19.1
Low, medium	High, very high	35.9	55.0	64.6
Low, medium, high	Very high	8.9	20.5	91.3

Note. Sensitivity = TP / (TP+FN), i.e. proportion of reoffenders who were above cutoff. Specificity = TN / (TN+FP), i.e. proportion of non-reoffenders who were below cutoff. TP = true positive; FN = false negative; FP = false positive; TN = true negative.

Table 10.9: Relative risk ratios for RM2000/s and OSP categories

Categories	RM2000/s		OSP	
	Ratio	(95% CI)	Ratio	(95% CI)
Medium vs. low	1.78	(1.13, 2.82)	2.22	(1.41, 3.50)
High vs. medium	1.62	(1.22, 2.16)	1.65	(1.23, 2.23)
Very high vs. high	1.82	(1.30, 2.55)	2.50	(1.82, 3.43)

10.4 Implications

Prediction of sexual reoffending

The results suggest that OSP has the potential to predict contact sexual reoffending in a more robust way than RM2000 (see also Chapter 13 for a summary of some further research conducted to validate and recalibrate OSP – this Chapter also notes how it has been incorporated into a new Risk of Serious Recidivism (RSR) tool which is being used within the Case Allocation System (National Offender Management System, 2014)). The development of OSP was grounded on the premise that focusing upon contact sexual offences would identify those current and past sex offenders most likely to cause serious harm, assisting with the public protection agenda. As well as improving sensitivity – identification of likely (contact) reoffenders – use of OSP would be likely to lead to greater proportions of sex offenders being classified as low risk. Both of these benefits would lead to more efficient resource allocation and therefore to more effective work to prevent sexual recidivism. It should nevertheless be stressed that, whatever risk assessment tool is used, prediction of contact sexual reoffending is a difficult task and only moderate predictive validity can be achieved.

Practical considerations if the new predictors are implemented

If OSP was to be brought into full operational use, several practical issues would need to be addressed. Notably, the introduction of a new risk prediction tool has implications for staff training and IT systems, and needs to be timed carefully. In NOMS, the introduction of OSP could be synchronised with other developments in offender assessment, including OASys IT improvements and updates to the OGRS and OASys operational predictors (see Chapters 8 and 9). The links between OSP and the OASys system for clinical rating of RoSH would also need to be considered (as they have been in the Case Allocation System). In the police service, the Violent and Sex Offender Register (ViSOR) IT system could be amended to allow OSP to be recorded. In both NOMS and the police, training or other learning support (e.g. provision of factsheets) for those dealing with sex offenders would be required, and operational guidance would be developed. While these activities would carry a cost in staff time, OSP should be quicker to administer, as unlike RM2000/s it requires no knowledge of the offender's past other than can be coded through review of a simple PNC criminal history printout. The 32 point scoring chart is very different to, but not necessarily more complex than, RM2000/s's two-stage scoring process, and experience from implementation of a similar system in OGP and OVP (Howard, 2009) suggests that assessors will not find it difficult.

At the time of publication, the necessary changes to NOMS IT systems were being considered to support the RSR tool into which OSP has been incorporated.

Specialisation in criminal careers

It is apparent that sex offenders specialise in particular types of sexual offence, and that patterns of reoffending and risk factors differ between the four types of sexual offence identified here. OSP has improved prediction of contact sexual reoffending by identifying the risk factors particularly associated with this outcome and not including those risk factors only associated with non-contact sexual reoffending. This information on specialisation and offence-specific risk factors may be relevant to treatment as well as assessment, and further consideration of the role of dynamic risk factors may be worthwhile, especially if combined with SARN clinical assessments.

Additional data analysis, not detailed here, has identified promising means of categorising sexual offenders' risks of paraphilia and indecent image reoffending, which again draw heavily upon specialisation in their selection of offence-specific risk factors. Appropriate guidance on risk factors for non-contact recidivism could thus be given to staff working with such offenders.

11. Criminogenic need measurements

While Chapters 8 to 10 focused upon predictors of differing types of reoffending, this chapter shifts attention to the OASys measurements of discrete criminogenic needs, ensuring adherence to the 'What Works' criminogenic need principle as well as the risk principle. Key points are as follows:

- The underlying factor structure of the scored OASys questions corresponds to the eight OASys criminogenic need sections.
- To maximise the item-scale correlations and the dynamism of the scales, as well as alignment to OGP2/OVP2, the analysis supports some changes to the questions which are scored – three questions being removed from the scoring and three being added.
- These amendments leave 31 scored questions across the eight criminogenic need scales, but all the scales (bar one – lifestyle and associates) now have four questions and a 0–8 scale.
- The revised scales were found to be independently associated with reoffending, with some changes required to the criminogenic need cut-off points.
- The above changes have an impact upon the criminogenic need prevalence rates across five of the scales (although a relatively small change for two of the scales) – adjustments in the allocation of resources would be required to ensure that interventions were available to address the revised criminogenic need levels.
- Fewer scales were found to be independently predictive for the Black, Asian and Minority Ethnic (BME) sub-groups, akin to the lower OGP1 predictive validity for BME offenders reported in Chapter 3.

11.1 Context

While the OASys predictors of general and violent reoffending (OGP/OVP – see Chapter 9) have improved the measurement of an offender's likelihood of reoffending, the focus in this chapter is upon evaluating the measurement of discrete criminogenic needs, ensuring adherence to the 'What Works' criminogenic need principle as well as the risk principle (McGuire, 1995). While the latter requires correspondence between the intensity of intervention and an offender's risk of reoffending, the former requires that, on the grounds of efficiency and effectiveness, the interventions should be targeted towards dynamic and changeable criminogenic needs. In other words, the risk principle helps to identify ***which offenders*** should receive the available interventions and the criminogenic need principle focuses on ***which problems*** should be addressed.

In its current format, the full OASys assessment scores eight 'criminogenic needs' using 31 practitioner-completed questions (see Appendix A). As shown by Table 11.1, the scales focus on either individual-level factors, in terms of 'internal' disposition, personality, reasoning and temperament, or 'external' social or societal factors and their influences on offending behaviour. The questions used within each section are totalled to produce overall section scores. For each of these sections, the offender is judged to have a need if the section score exceeds a designated cut-off point. These cut-off points were set no lower than 2+ (equivalent to significant problems on one question or some problems across two questions) and were calculated using previously combined OASys/reoffending data – offenders with scores above the cut-off points were found to have above average reoffending rates, indicating that interventions to address the need may be beneficial in terms of reducing reoffending (see Moore, 2009b, for a previous example of this approach). Importantly, the section scores that indicate a need for intervention are different in the individual sections. For example, a score of **2** on accommodation indicates a need for intervention, whereas in the alcohol misuse section, a score of **4** indicates a need for intervention.

Table 11.1: Scored OASys scales

Criminogenic need	Scored questions	Scale range	Cut-off
Accommodation	3.3, 3.4, 3.5, 3.6	0–8	2+
ETE	4.2, 4.3, 4.4, 4.5	0–8	3+
Relationships	6.1, 6.3, 6.6	0–6	2+
Lifestyle & associates	7.2, 7.3, 7.5	0–6	2+
Drug misuse	8.4, 8.5, 8.6, 8.8, 8.9	0–10	2+
Alcohol misuse	9.1, 9.2, 9.3, 9.5	0–8	4+
Thinking & behaviour	11.5, 11.6, 11.7, 11.9	0–8	4+
Attitudes	12.1, 12.4, 12.5, 12.8	0–8	2+

11.2 Approach

Research questions

To guide the analysis, the following four research questions were set:

1. Do the scored questions within each scale measure discrete individual-level or social characteristics? (Internal reliability)
2. What are the common factors underlying the scored questions? (Construct validity)
3. Are the OASys scales measuring criminogenic needs?
4. What are the optimum criminogenic need cut-off points?

These questions have been addressed previously (Moore, 2009b) but on older samples and it was acknowledged that the analysis should be repeated once larger samples were available for the age, gender and ethnicity sub-groups. This was deemed important, bearing in mind that previous research has found that criminogenic needs and their associations with offending differ between males and females and between different age groups (e.g. Hollin and Palmer, 2006; Raynor, 2007).

The samples

Assessments were restricted to those completed at the Pre-Sentence Report stage or at the start of either a community order or the licence period of a custodial sentence between April 2005 and March 2008 inclusive. It was ensured that these assessments included all 31 scored criminogenic need questions from the core assessment, a sentence date and consistent risk of serious harm data. The assessments were then submitted to the Police National Computer (PNC) research database managed within the Ministry of Justice (MoJ).[189] Once successfully matched, the PNC records were processed to determine whether the cases could be followed up for 24 months at liberty, taking into account periods spent in custody and allowing three months for sentence and data entry to occur. This left a final sample size of 180,687 cases for use in the analysis.

Analysis

The most common ***internal reliability*** measure is Cronbach's alpha, and this was used to measure how well the individual questions in each OASys scale correlated with the sum of the remaining questions. Alpha scores generally increase when the correlations between the questions increase, thus indicating the extent to which each set of questions can be treated as measuring a discrete characteristic, i.e. an individual-level or social problem. While a lenient cut-off of 0.6 can be used in exploratory research, many researchers require a cut-off of 0.7 for a scale to be considered 'adequate' and 0.8 for a scale to be considered 'good'.[190]

By comparing each scale's overall alpha score to the score produced when each individual question was removed, the results were used to indicate which questions were not contributing to the scale's internal reliability. Item-scale correlations were also calculated to demonstrate which questions were poorly correlated with the total of scores on all other items.

While Cronbach's alpha was used to verify the internal reliability of each scale, factor analysis was used to measure the overall ***construct validity*** of the core OASys assessment.[191] Factor analysis assesses the interrelationships among a large number of questions and then explains these questions in terms of their common underlying dimensions (factors). The makeup of these factors thus gives information on the relationships between the individual questions. By comparing the factors to the established scales, the results can be used to:

189 PNC numbers were recorded within OASys for most offenders, and an automatic matching procedure found reliable PNC numbers for most of the remaining cases. Cases in which the PNC did not record the offender's sex or recorded an unfeasible date of first or current conviction were rejected.

190 Oppenheim states that 'Reliability, or self-consistency, is never perfect; it is always a matter of degree' (1996:159). The required standard of reliability also varies between subject areas. For example, cognitive tests tend to be more reliable than tests of attitudes or personality. More specifically, it is easier to construct a reliable test of a particular attitude than of a general one.

191 The analysis used the principal components method, producing uncorrelated factors, as well as varimax rotation to maximise the variance of the loadings, helping to link each question to a single factor. The factors were restricted to those with an eigenvalue greater than one, recognising that further factors were contributing little to the explanation of variance in the variables.

(i) validate the established scales by demonstrating that their constituent questions load on the same factors;

(ii) propose the construction of new scales; and/or

(iii) propose the removal of questions which are weakly correlated with any specific factor and instead cross-load across factors (as indicated by low factor loadings, e.g. less than 0.4).

In order to assess which scales were measuring not only distinct problem areas but independent criminogenic needs, logistic regression was used to look at the associations with reoffending, taking into account the correlations between the scales themselves. Odds ratios were used to establish criminogenic need cut-off points, comparing the odds of reoffending for offenders with a particular score to the average odds of reoffending.[192]

11.3 Results

Internal reliability

In addition to the 31 scored questions currently included in the criminogenic need scales, the following items were incorporated into the analysis as these are included within the second iterations of OGP and OVP (see Chapter 9): 6.4 (Relationship with partner), 6.7 (Domestic violence: Perpetrator and/or victim), 11.2 (Impulsivity) and 11.4 (Temper control).

As shown by Table 11.2, the internal reliability alpha scores for the criminogenic need scales were as follows:

- Accommodation (0.93);
- ETE (0.79);
- Relationships (0.67);
- Lifestyle and associates (0.66);
- Drug misuse (0.81);
- Alcohol misuse (0.85);
- Thinking and behaviour (0.77); and
- Attitudes (0.70)

Six of the eight scales had adequate or good reliability (with Crohnbach's alpha scores of 0.7 or above), demonstrating that the questions within these scales were measuring discrete individual-level or social characteristics. While the two remaining scales (relationships and lifestyle/associates) had Cronbach's alpha scores below 0.7, both were above 0.65.

[192] Odds ratios can be used to compare whether the probability of a certain event is the same for two groups: an odds ratio of 1 indicating that the event is equally likely in both groups; an odds ratio greater than one indicating that the event is more likely in the first group; and an odds ratio less than one indicating that the event is less likely in the first group.

Table 11.2 also sets out the Cronbach's alpha scores for the scales when each individual question was removed (where the cell is empty there is no such scored question within the relevant section). Most scores were lower than the overall alpha, suggesting that the numbered question was contributing to the measurement of a discrete characteristic. For example, when question 9.1 was removed, the section's alpha fell from 0.85 to 0.77. For two of the 35 questions (indicated in **bold** in the table), the section's alpha score increased when it was removed, suggesting that it was not contributing to the measurement of that characteristic. These two questions were as follows:

- 4.2: Is the person unemployed or will be unemployed on release?
- 11.4: Temper control

Turning to the item-scale correlations of the scored OASys questions, the following three questions (indicated in **bold** in Table 11.3) had correlations below 0.4, two of which are from the relationships section (which, as indicated above, had an alpha score below 0.7):

- 6.3: Experience of childhood;
- 6.4: Current relationship with partner; and
- 11.4: Temper control

As can be seen, 11.4 (one of the questions used within OGP2/OVOP2 but not the current criminogenic need scales) had an item-scale correlation below 0.3 and lowered the section's alpha score. The decision was thus made to remove this question from the further analysis.

Table 11.2: Internal reliability of current OASys scales

Scale	Cronbach's alpha	Cronbach's alpha for the scale when numbered question is deleted									
		1	2	3	4	5	6	7	8	9	10
Accommodation	. 933			.906	.907	.916	.922				
ETE	. 787		**.825**	.669	.697	.747					
Relationships	. 668	.624		.642	.631		.581	.603			
Lifestyle & associates	. 663		.506	.555		.636					
Drug misuse	. 807				.768	.760	.795		.785	.732	
Alcohol misuse	. 851	.771	.781	.830		.848					
Thinking & behaviour	. 770		.751		**.789**	.714	.713	.720		.726	
Attitudes	.699	.612			.639	.605			.683		

Key: **Bold font** = increased alpha score for section when the question is removed.

Table 11.3: Item-scale correlations of current OASys scored questions

Scale	Item-scale correlation for numbered question									
	1	2	3	4	5	6	7	8	9	10
Accommodation			.867	.860	.832	.815				
ETE		.470	.725	.670	.591					
Relationships	.408		**.366**	**.390**		.499	.455			
Lifestyle & associates		.519	.485		.423					
Drug misuse				.612	.647	.515		.555	.726	
Alcohol misuse	.781	.758	.653		.612					
Thinking & behaviour		.460		**.318**	.606	.608	.578		.555	
Attitudes	.520			.480	.539			.406		

Key: **Bold font** = item-scale correlation below 0.3.

Construct validity

The results of the factor analysis revealed eight factors underlying the 34 remaining questions, explaining 62 per cent of the variation in the variables. These eight factors correspond to the eight OASys sections which are currently scored as criminogenic needs. The loadings listed in the final column of Table 11.4 indicate the correlation between each OASys question and its respective factor – the higher the loading, the greater the contribution to the factor. All questions had a factor loading in excess of 0.4. Two of the 34 questions (indicated in bold in Table 11.4) did not fall into the factors corresponding to the sections within which they reside. In both instances, there was a clear cross-loading with the correct sections: question 11.9 had a factor loading of 0.47 for the thinking and behaviour factor and question 7.2 had a factor loading of 0.39 for the lifestyle and associates factor.

Table 11.4: Underlying factors of current scored questions

Factor (Variance explained)	Question	Loading
1. Accommodation (10.0%)	3.3: Currently of no fixed abode or in transient accommodation	.923
	3.4: Suitability of accommodation	.891
	3.5: Permanence of accommodation	.875
	3.6: Suitability of location of accommodation	.855
2. Drug misuse (9.6%)	**7.2: Regular activities encourage offending**	**.417**
	8.4: Current drug noted	.760
	8.5: Level of injected drugs	.769
	8.6: Ever injected drugs	.647
	8.8: Motivation to tackle drug misuse	.647
	8.9: Drug use and obtaining drugs a major activity/occupation	.813
3. Alcohol misuse (8.5%)	9.1: Is current use a problem?	.885
	9.2: Binge drinking or excessive use of alcohol in last month	.874
	9.3: Frequency and level of alcohol misuse in the past	.774
	9.5: Motivation to tackle alcohol misuse	.748
4. Employment (7.8%)	4.2: Is the person unemployed, or will be unemployed on release	.621
	4.3: Employment history	.804
	4.4: Work-related skills	.811
	4.5: Attitude to employment	.692
5. Thinking & behaviour (7.6%)	11.2: Impulsivity	.475
	11.5: Ability to recognise problems	.760
	11.6: Problem-solving skills	.760
	11.7: Awareness of consequences	.783
6. Attitudes (7.4%)	**11.9: Understands other people's views**	**.518**
	12.1: Pro-criminal attitudes	.641
	12.4: Attitude towards supervision/licence	.574
	12.5: Attitude to community/society	.703
	12:8: Motivation	.646

Factor (Variance explained)	Question	Loading
7. Relationships (6.4%)	6.1: Current relationship with close family members	.518
	6.3: Experience of childhood	.543
	6.4: Current relationship with partner	.596
	6.6: Previous experience of close relationships	.723
	6.7: Domestic violence: Perpetrator and/or victim	.730
8. Lifestyle & associates (5.1%)	7.3: Easily influenced by criminal associates	.507
	7.5: Recklessness and risk-taking behaviour	.598

Revising the OASys scales

Reviewing the above findings on the internal reliability of the scales and the overall construct validity of the assessment, alongside the need to establish concise scales which are as dynamic as possible, the decision was made to remove three further items from the scales. The reasoning was as follows:

- 6.3 'Experience of childhood' – this had the lowest item-scale correlation within the section (below 0.4) and is a non-dynamic question due to its focus on earlier childhood and adolescence.
- 8.6 'Ever injected drugs' – this had the lowest item-scale correlation within the section and is not fully dynamic due to its consideration of previous as well as current injection.
- 11.9 'Understands other people's views' – this fell into the incorrect factor and had a relatively low item-scale correlation within its section.

These amendments left 31 scored questions across the eight criminogenic need scales. While the overall number of questions is unchanged from the current system, seven of the scales now have four questions and a 0–8 scale – the exception being the lifestyle and associates section which still has three questions and a 0–6 score range. Three questions have been removed from the scoring (6.3, 8.6 and 11.9), with three being added (6.4, 6.7 and 11.2).[193]

These changes were found to have little impact upon the scales' internal reliability scores, with six of the scales still having Cronbach's alpha scores of at least 0.7 and the remaining two scales (relationships and lifestyle/associates) having an alpha score above 0.65. Checking the alpha scores for differing offender sub-groups revealed some variation, with the alpha score for the attitudes scale falling below 0.7 for older offenders (those aged 25+) and those of Black, Asian and Other ethnicity.

Re-running the factor analysis on the revised 31 questions, all questions fell into the factors corresponding to the sections within which they reside with the exception of 11.2 (impulsivity) which

[193] These changes align to the questions used in OGP2 and OVP2 – the former three questions are not used in these predictors, while the latter three are used.

cross-loaded across the thinking/behaviour and lifestyle/associates sections. Running the factor analysis for the different offender sub-groups revealed a few further instances of cross-loading:

- For 18–20 year olds, question 6.1 (current relationship with close family members) fell into the accommodation section.
- For 25–40 year olds, question 7.2 (regular activities encourage offending) fell into the drug misuse section.
- For female and Asian offenders, question 7.5 (recklessness and risk taking behaviour) fell into the thinking and behaviour section.
- For Asian offenders, the lifestyle/associates and drug misuse sections were merged together into one factor.

Are the OASys scales measuring criminogenic needs?

In the development of the risk-needs model, Andrews and Bonta (1995 :176) stressed the importance of distinguishing between criminogenic needs and more general needs according to their relationship with reoffending, stating that the former are 'the dynamic attributes of an offender that, when changed, are associated with changes in the probability of recidivism'. Criminogenic needs can thus be defined as those individual risk factors which contribute to or are supportive of offending and which are amenable to change.

In order to assess which of the OASys scales were measuring independent criminogenic needs, logistic regression was used to take into account the correlations between the scales themselves. The OGRS 3 score (based upon static criminal history and offender demographic factors) and all eight scales were entered into the model. As shown by Table 11.5, all eight scales were found to have an independently significant association with reoffending. In other words, all scales were measuring distinct problem areas and independently significant criminogenic needs. Across all the scales, the odds ratios was greater than one, indicating that the odds of reoffending for those with higher scores were greater than the odds of reoffending for those with lower scores.

Table 11.5: Associations with 24-month reoffending

Scale	Parameter estimate	Standard error of estimate	Significance	Odds ratio
OGRS 3	.039	.000	***	1.040
Accommodation	.026	.002	***	1.027
ETE	.043	.003	***	1.044
Relationships	.012	.003	***	1.012
Lifestyle & associates	.031	.004	***	1.032
Drug misuse	.064	.003	***	1.066
Alcohol misuse	.039	.002	***	1.040
Thinking & behaviour	.009	.003	**	1.009
Attitudes	.040	.004	***	1.041
Constant	*-2.673*	*.018*	***	*.069*

Asterisks indicate whether associations are significant (*** p<.001; ** p<.01).

Re-running the regression model for the differing offender sub-groups revealed that thinking and behaviour was the section most commonly not predictive. Fewer scales were found to be independently predictive for the Black, Asian and Minority Ethnic (BME) sub-groups – the predictive scales for Black offenders were accommodation, ETE, drug misuse and alcohol misuse, whilst the predictive scales for Asian offenders were accommodation, drug misuse and alcohol misuse.

Revising the criminogenic need cut-off points

Odds ratios were used to set appropriate criminogenic need cut-off points for the eight scales, comparing the odds of reoffending for offenders with a particular score to the average odds of reoffending. Of the 180,687 offenders in the OASys/PNC sample, 49.8% had reoffended within two years. As shown by Table 11.6, there was a point across all eight scales at which the odds ratio increased to a value greater than one, i.e. where the reoffending rate surpassed 49.8%, so that the odds of reoffending for an individual with that score exceeded the average odds of reoffending. This was the point at which offenders were judged to have a criminogenic need, adjusted to maintain a minimum cut-off point of 2+ (equivalent to significant problems on one question or some problems across two questions). For example, in relation to the attitudes scale, the reoffending rate increased from 32.6% for a score of 0 to 77.5% for a score of 8. The odds ratio increased to a value above one for those offenders who scored 2 with an above average reoffending rate of 52.5%. On this basis, those offenders with scores from 0 to 1 were judged to have no need, while those with scores from 2 to 8 were judged to have a need.

Table 11.6: 24-month reoffending rates by revised OASys scales

Score	24-month reoffending rate by scale							
	Accomm	ETE	Relationships	Lifestyles & associates	Drug misuse	Alcohol misuse	Thinking & behaviour	Attitudes
0	41.7%	29.1%	42.2%	30.9%	38.7%	45.9%	30.4%	32.6%
1	50.9%	41.0%	52.3%	40.5%	59.2%	46.2%	35.0%	42.6%
2	55.3%	40.3%	54.1%	49.4%	58.7%	48.2%	37.7%	52.5%
3	62.0%	49.3%	54.1%	60.8%	63.0%	48.4%	42.6%	61.8%
4	61.6%	55.5%	49.7%	65.9%	64.0%	50.5%	51.0%	68.7%
5	64.5%	61.3%	51.0%	73.3%	67.1%	53.3%	56.7%	71.8%
6	64.0%	66.0%	50.7%	76.3%	71.6%	55.5%	57.8%	74.2%
7	100.0%	70.1%	54.7%		77.8%	60.5%	62.7%	76.2%
8	62.2%	71.7%	55.6%		79.8%	65.1%	66.6%	77.5%

Key: Criminogenic need level: No need Need

Table 11.7 sets out the differences between the current criminogenic need scales and the revised scales. As can be seen, the cut-off points for the ETE and lifestyle/associates sections have increased by one point, reducing the proportion of offenders scored as having these needs (by 10 and 20 percentage points respectively). The criminogenic need prevalence rates have also changed across the relationships, drug misuse and thinking/behaviour sections due to the changes in the scored questions (most notably for the thinking/behaviour section which increased by eight percentage points).

Table 11.7: Current and revised criminogenic need prevalence rates

OASys section	Current			Revised		
	Scale	**Cut-off**	**% with need**	**Scale**	**Cut-off**	**% with need**
Accommodation	0–8	2+	36.1	0–8	2+	36.1
ETE	0–8	3+	57.8	0–8	4+	47.6
Relationships	0–6	2+	54.1	0–8	2+	56.5
Lifestyle & associates	0–6	2+	57.7	0–6	3+	37.5
Drug misuse	0–10	2+	38.2	0–8	2+	37.7
Alcohol misuse	0–8	4+	34.9	0–8	4+	34.9
Thinking & behaviour	0–8	4+	54.6	0–8	4+	62.9
Attitudes	0–8	2+	49.6	0–8	2+	49.6

Comparing the odds of reoffending for offenders with a particular score to the average odds of reoffending for different age, gender and ethnicity sub-groups revealed some variation in the optimum cut-off points. As Table 11.8 demonstrates, females had a higher cut-off point than males for the ETE, relationships and alcohol misuse sections and a lower cut-off point for the lifestyle/associates section. Comparing the youngest (18–20) and oldest (41+) offenders, the latter had a lower cut-off point for the ETE, lifestyle/associates and thinking/behaviour sections and a higher cut-off point for the alcohol misuse section. There were also some differences by ethnicity, with Asian offenders having lower cut-off points than White offenders for the ETE, lifestyle/associates and alcohol misuse sections.

Table 11.8: Criminogenic need cut-off points for revised OASys scales by gender, age and ethnicity

	n	24-month reoffending rate	Criminogenic need cut-off point by scale							
			Accom	ETE	Relationship	Lifestyle	Drugs	Alcohol	Thinking	Attitudes
All	180,687	49.8%	2+	4+	2+	3+	2+	4+	4+	2+
Male	157,158	50.9%	2+	3+	2+	3+	2+	4+	4+	2+
Females	23,529	41.9%	2+	4+	3+	2+	2+	5+	4+	2+
18–20	31,948	63.3%	2+	4+	2+	3+	2+	4+	5+	2+
21–24	32,995	56.1%	2+	3+	2+	3+	2+	4+	4+	2+
25–40	83,801	49.7%	2+	3+	2+	2+	2+	5+	4+	2+
41+	31,941	29.8%	2+	3+	2+	2+	2+	5+	3+	2+
White	139,729	51.8%	2+	4+	2+	3+	2+	4+	4+	2+
Black	8,743	50.7%	2+	4+	2+	3+	2+	2+	4+	2+
Asian	6,487	43.9%	2+	3+	2+	2+	2+	2+	4+	2+
Mixed	3,655	57.0%	2+	4+	2+	3+	2+	3+	4+	2+
Other	941	38.6%	2+	4+	2+	2+	2+	2+	3+	3+
Male 18–20	28,401	65.1%	2+	4+	2+	3+	2+	4+	5+	2+
Male 21–24	29,071	57.4%	2+	3+	2+	3+	2+	4+	4+	2+
Male 25–40	72,182	50.6%	2+	3+	2+	3+	2+	5+	4+	2+
Male 41+	27,502	30.4%	2+	3+	2+	2+	2+	5+	4+	2+
Female 18–20	3,547	49.2%	2+	5+	2+	3+	2+	4+	5+	3+
Female 21–24	3,924	46.6%	2+	5+	3+	3+	2+	4+	4+	2+
Female 25–40	11,619	44.2%	2+	4+	3+	2+	2+	2+	4+	2+
Female 41+	4,439	26.2%	2+	4+	3+	2+	2+	5+	4+	2+

Key: Below cut-off for whole sample Above cut-off point for whole sample

11.4 Implications

The analysis revealed that the underlying factor structure of the scored OASys questions corresponds to the eight OASys criminogenic need sections. To maximise the item-scale correlations and the dynamism of the scales, as well as alignment to OGP2/OVP2, the analysis supports some changes to the questions which are scored. When implementing these changes, the revised scales were found to be independently associated with reoffending, with some changes required to the criminogenic need cut-off points. Fewer scales were found to be independently predictive for the BME sub-groups, akin to the lower OGP1 predictive validity for BME offenders reported in Chapter 3.[194]

More specifically, the implications for sections 3 to 12 of OASys are as follows:

- The amendments set out in this chapter leave 31 scored questions across eight criminogenic need scales, but all the scales (bar one – lifestyle and associates) now have four questions and a 0–8 scale.
- Three questions have been removed from the scoring (6.3, 8.6 and 11.9), with three being added (6.4, 6.7 and 11.2).
- The criminogenic need cut-off points for the ETE and lifestyle/associates sections have increased by one point.
- These changes have an impact upon the criminogenic need prevalence rates across five of the scales (although a relatively small change for two of the scales) – adjustments in the allocation of resources would be required to ensure that interventions were available to address the revised criminogenic need levels.

[194] The analyses conducted within the Chapter 3 study also indicated that there was limited scope to improve the validity of the OASys-based predictive scores for BME offenders by building separate scores for each ethnic group.

12. Systematic review of factors related to general and violent reoffending

This chapter presents a systematic review of the literature on the dynamic risk and protective factors for general and violent reoffending, recognising that OASys must not only continue to pass stringent reliability/validity performance criteria but must also continue to reflect the research literature on which it is based. Thirty-two UK and international studies published between January 2000 and November 2011 (heterogeneous in terms of populations, methodology and data reporting) were included in the review. Key points are as follows:

- No new risk domains were identified that would be worthwhile additions to OASys.
- In terms of more specific items within the domains, not all items were consistently identified and those that were most consistently identified matched closely to specific OASys questions.
- Gang membership, which is not currently recorded within OASys, was found to be predictive of future violent reoffending in one relatively large US study. Consideration could be given to including a question on gang associations/activities within the current lifestyle and associates section.
- Further reviews of the literature could be undertaken using the same systematic approach, helping to ensure that offender assessment policy within NOMS continues to reflect the most up-to-date knowledge about risk and protective factors.
- There is a clear need for further studies identifying: (i) positive factors which are negatively correlated with reoffending as well as those which moderate the impact of specific risk factors; and (ii) whether there are differences between the dynamic risk and protective factors according to age, gender and ethnicity. Further attention also needs to be given to which dynamic factors are truly causal, where changes over time are associated with changes in future offending behaviour when other factors are held constant.

12.1 Context

If OASys is to maintain its central role in assessing risk and need, it must not only continue to pass stringent performance criteria relating to its current content, but must also continue to reflect the wider research literature on which it is based. A systematic review was thus undertaken to ensure that any forthcoming changes to OASys reflect the most up-to-date knowledge about risk and protective factors, and take into account developments within the research literature since OASys was initially developed. The review ensures that any proposed changes will be informed by recent reviews and meta-analyses of the predictive validity of risk assessment tools (Yang *et al*., 2010), developments within the field such as the evaluation of fourth generation tools (Brennan, Dieterich and Ethret, 2009), and the inclusion of temporal adjustments to risk assessment (Harris and Rice, 2007). Further recent developments in the What Works literature include the focus on social capital as a protective factor

(Hochstetler, DeLisi and Pratt, 2010); exploring long-term pathways into crime using longitudinal studies (Farrington, Ttofi and Coid, 2009), and a growing body of evidence on the application of risk assessment tools to female (Smith *et al*., 2009) and ethnic minority (Fass *et al*., 2008) offenders.

12.2 Approach

Research questions

The aim of the systematic review was to examine the evidence base relating to the dynamic risk (stable and acute) and protective factors for general and violent reoffending among adult offenders. The focus was upon 'dynamic' factors – those that are linked to reoffending and which are amenable to change, such as antisocial cognitions, values, and attitudes – as opposed to 'static' factors which are typically historical and are not amenable to change, such as gender, age, and family background (Gendreau, Little and Goggin, 1996). Protective factors were defined broadly to cover positive factors negatively correlated with reoffending, not only those moderating the impact of specific risk factors.[195]

The study had three key objectives:

- To critically evaluate research relating to the dynamic risk (stable and acute) and protective factors for general reoffending among adult offenders.
- To critically evaluate research relating to the dynamic risk (stable and acute) and protective factors for violent reoffending among adult offenders.
- To identify potential improvements to OASys.

The three objectives were broken down into the following questions and answered through a systematic review of the evidence.

Question 1: What are the dynamic risk (stable and acute) and protective factors for general reoffending among adult offenders?

- What are the dynamic risk factors relating to general reoffending?
- What are the positive factors which are negatively correlated with general reoffending or which moderate the impact of specific risk factors?
- Which of the identified risk and protective factors have the strongest correlations with general reoffending?
- Are the dynamic risk and protective factors for general reoffending consistent for adult offenders of different age, gender and ethnicity?

Question 2: What are the dynamic risk (stable and acute) and protective factors for violent reoffending among adult offenders?

- What are the dynamic risk factors relating to violent reoffending?

[195] As set out in Chapter 6, the former are sometimes termed 'promotive' rather than 'protective'.

- What are the positive factors which are negatively correlated with violent reoffending or which moderate the impact of specific risk factors?
- Which of the identified risk and protective factors have the strongest correlations with violent reoffending?
- Are the dynamic risk and protective factors for violent reoffending consistent for adult offenders of different age, gender and ethnicity?

Question 3: Which factors identified through the systematic review (questions 1 and 2 above) are not currently within OASys and could be beneficial in predicting general and violent reoffending?

The focus of the final question upon items not currently in OASys recognises the ability to test and validate the current content of the tool through combined OASys and PNC data (as demonstrated in the previous chapters in this compendium).

Procedure

Systematic searching

Four relevant databases using the EBSCO Discovery Service (EDS) were accessed for this review:

- Criminal Justice Abstracts
- Medline
- PsycInfo
- SocIndex

The following initial search terms were devised to identify the most relevant studies for the research questions:[196]

(1) TX (offend* OR reoffend* OR recidivis* OR crime* OR criminal* OR violen*) NOT TX (sexual* NOT sex* NOT adolescent* NOT partner* NOT juvenile* NOT child* NOT "mental illness" NOT schizophrenia* NOT domestic*)

AND

(2) TX predict* OR promot* OR protect* OR dynamic* OR stable* OR desist* OR caus* OR acute* OR positive*

AND

(3) TX risk* OR factor*

AND

(4) TX assess*

[196] TX indicates that all text (title, abstract etc) was searched.

As indicated by these search terms, studies which focused on domestic or sexual violence were excluded.[197] The search was further restricted to studies reported in English and published between January 2000 and November 2011 (bearing in mind that the initial development work for OASys was completed by 2001).

The following websites were also searched for relevant, unpublished data, or links to relevant citations:

- Ministry of Justice (MoJ) and Home Office websites (www.justice.gov.uk; www.homeoffice.gov.uk)
- National Institute of Justice (www.nij.gov)
- American Correctional Association (www.aca.org)
- Australian Institute of Criminology (www.aic.gov.au)
- Correctional Service Canada (www.csc-scc.gc.ca)
- National Institute of Corrections (www.nicic.gov)
- OpenSIGLE (for grey literature; www.greynet.org)

To identify further studies not identified by the database search and to fill any gaps in the evidence after full text screening, these sources were also searched:

- National Criminal Justice Reference Service (NCJRS), Google Scholar, Cochrane Collaborations
- Citations from included references and meta analyses
- Studies that cite the included references, identified using citation databases
- Authors of included studies
- Tables of contents of key journals in the field for the last five years of publication

The review did not extend to searching for non-published evidence through contacting relevant experts.

Screening and quality assessment

Abstracts were screened independently by two reviewers initially, and disagreements were resolved by discussion and consensus.[198] Inter-rater reliability was monitored after 10% of abstracts were screened.[199] The review continued with a single reviewer screening each abstract to enable the project deadlines to be met. The full screening inclusion criteria are set out in Table K1 in Appendix K. As indicated, systematic reviews were excluded per se as not reporting primary empirical data, but

[197] Research has previously examined dynamic risk factors in relation to sexual recidivism (see, for example, Mann, Hanson and Thornton (2010). It is also known that there are differing specific risk factors for domestic violence (see, for example, the Spousal Assault Risk Assessment Guide (SARA): Kropp, Hart, Webster, and Eaves, 1995).

[198] A third reviewer could have been used to broker any disagreements, but this was deemed impractical due to the need to review a large number of abstracts relatively quickly.

[199] The reliability score (Cohen's kappa) was required to be above $\kappa = 0.6$.

were used to identify additional primary studies as part of citation chasing. Where there was uncertainty about the relevance of a research report from the abstract, the full text was retrieved where possible.

Full texts were retrieved using Google Search engine, the British Library and the MoJ library, with studies being excluded on full text screening using the same criteria as for abstract screening. The first 10% of full text references were screened by two reviewers independently, to assess adequate reliability, and the remaining full text papers were screened by one researcher each.

Studies meeting the inclusion criteria were then assessed for their overall quality, with a quality assessment checklist being devised to review the full texts (Tables K2 and K3 in Appendix K). Each study was rated as high, mid or low quality across the following four dimensions:

- Reporting transparency
- Appropriateness of study design
- Quality of execution
- Relevance

Studies were then given an overall quality rating that equalled their lowest rating across the four dimensions. It was felt that (i) a study was only as good as the lowest of its ratings and (ii) this approach would minimise bias in determining overall quality. Details of the quality assessment ratings for each study are given in Appendix K.

Quality assessment for all included studies was carried out by one reviewer and checked by another, with a minimum of 10% of studies independently quality-assessed by two reviewers, and any differences resolved by discussion and reference to a third reviewer where necessary. Studies that were of low quality were then excluded so that the review's findings and conclusions would be based only on higher quality evidence.

Flow of literature

A flow of literature diagram is reported in Appendix L. The database searches resulted in the identification of 9,187 references, of which 2,503 were duplicates across the databases. A total of 6,684 unique studies and reports on abstract were screened, resulting in the exclusion of 6,466 studies. The agreed list of websites was searched for grey literature, resulting in the identification of a further 29 studies and reports. The searches of contents pages of relevant journals as well as citation chasing from reviews and meta-analyses returned an additional ten references.

A total of 36 full texts could not be obtained, leaving 221 full texts for further screening. Of these, 188 papers were then excluded. Most of these were excluded for not reporting on dynamic risk or protective factors relating to adult offenders for general or violent reoffending (125 studies).
Others were excluded for not reporting reoffending outcomes (25 studies), not reporting empirical data

(21 studies), being a systematic review (six studies, which were used for citation chasing), being published before 2000 (three studies) and for being assessed as low quality (nine studies). This left a final sample of 32 studies.

The studies

Of the 32 studies included in the review, 25 reported on dynamic risk factors associated with general reoffending and 13 reported on factors associated with violent reoffending (with six of the studies reporting on both). Twenty five of the studies were international studies, with the majority originating from the USA, Canada or Australia,[200] and the transferability of the findings from these studies to UK offenders cannot of course be guaranteed. The remaining seven studies were undertaken within the UK. As the studies were heterogeneous in terms of populations, methodology (e.g. some samples were comprised of volunteers), data analysis (e.g. use of differing statistical tests) and reporting, no meta-analysis or quantitative synthesis of the data was deemed possible. Consequently, identifying which risk and protective factors had the strongest correlations with general and violent reoffending was not straightforward.

Looking across the studies, it became apparent that there was insufficient evidence to assess whether the dynamic risk and protective factors for general and violent reoffending were consistent for adult offenders of different ages, gender and ethnicity. Three of the studies focused solely upon female offenders, but very few had mixed samples and explored whether the factors were common or unique to certain sub-groups. Manchak, Douglas and Siranosian (2009) was one study which did explore differences between male and female violent offenders, but the female sub-sample was small; 70 offenders compared to 1,035 male offenders.

A brief description of the 32 studies (which are listed separately in the References section following this chapter) is provided below. Consistent information was extracted for each study using the data extraction template set out in Appendix M, covering the specific measures used (predictors and outcome measures), the targeting of the assessment tools or interventions, and the final sample sizes.

The studies are grouped below into those which reported on:

- general and violent reoffending outcomes;
- general reoffending outcomes only; and
- violent reoffending outcomes only.

[200] One study was from Finland.

Within these groups, the studies are then listed according to: (i) UK versus international studies (UK studies listed first); and (ii) date (the most recent studies first). Within the study descriptions, references are made to various risk assessments instruments and other tools/scales – these are summarised in Appendix N. As can be seen, the sample sizes and the follow-up periods (where specified) varied greatly across the studies – some of the studies were based upon relatively small samples. It is also clear from the tables presented in this chapter that there was great variance in the extent to which other risk factors were controlled for in the analyses.

General and violent reoffending outcomes

UK studies

Craig, Browne, Beech and Stringer (2004) examined personality characteristics associated with general and violent reconviction, using the Special Hospitals Assessment of Personality and Socialization (SHAPS) scale. They assessed a sample of 121 convicted male violent and sexual offenders who were followed up over a period of two, five and ten years. Reconviction rates were calculated using data from the Home Office Offenders Index. The accuracy of the SHAPS scale scores were compared with those of actuarial measures (Static-99 and Risk Matrix 2000) in predicting reconviction.

International studies

Listwan, Van Voorhis and Ritchey (2007) explored the relationship between personality and long-term recidivism, using a cohort 277 federal prisoners in Indiana (USA). Data were derived from interviews, pre-sentence investigation reviews and the Jesness Inventory. Recidivism data during a period of 10 to 12 years were collected through the use of the National Crime Information Centre database, with the offence categories collapsed to include drug, property and violent offences. Risk of reoffending was measured using a modified version of the Salient Factor score.

Wormith, Olver, Stevenson and Girard (2007) examined the prediction of general and violent recidivism over a ten year follow-up period using the (i) Hare Psychopathy Checklist – Revised (PCL-R), (ii) the Level of Service/Case Management Inventory (LS/CMI) and (iii) DSM-III Antisocial Personality Disorder (APD) criteria on a sample of 61 Canadian adult male offenders. An interview protocol was prepared to enable the scoring on all three tools. The Canadian Police Information Centre (CPIC) was accessed to obtain and code recidivism data.

Siddiqi (2006) examined the likelihood of pre-trial rearrest for violent felony offences and rearrest in general, as well as the risk of pre-trial failure among New York City (US) defendants. The analysis was conducted in two phases. Data were drawn from a cohort of arrests made between January and March 2001, in which defendants were prosecuted on new charges. In Phase I, the likelihood of a rearrest being for a violent felony offence was examined in a sample of 4,827 defendants. It was determined whether the same variables predicted pre-trial rearrest in general by looking at the full at-risk sample (n =27,630).

Mills, Kroner and Hammati (2004) examined the validity of the Measures of Criminal Attitudes and Associates (MCAA) in the prediction of general and violent recidivism. A volunteer sample of 144 Canadian adult male offenders with a sentence of two or more years was followed up for one year after their release from custody using offender files and official police records.

Piquero, Brame, Mazerolla and Haapanen (2001) used a prospective longitudinal dataset of 524 male offenders released from the California (US) Youth Authority. The offenders were followed up over a seven-year post parole period to examine the effect of life circumstances, i.e. marriage, employment, drug use, alcohol use and street time, on general and violent offending.

General reoffending outcomes only

UK studies

Farrington, Ttofi and Coid (2009) investigated the life success at ages 32 and 48 of four categories of males: non offenders, adolescence-limited offenders (convicted only at ages 10–20), late-onset offenders (convicted only at ages 21–50) and persistent offenders (convicted at both ages 10–20 and 21–50) using data from the Cambridge Study in Delinquent Development, a survey of 411 South London males. The subjects were followed up from age eight to 48 in repeated personal interviews. The aim was to establish the most important variables for predicting the future offending careers of offenders and non-offenders at differing ages, including an identification of the main risk factors for predicting persistent offenders (n=70) compared to adolescence-limited offenders (n=53).

May, Sharma and Stewart (2008) drew on three resettlement surveys of UK prisoners conducted in 2001, 2003 and 2004. Data from the three surveys were matched with PNC reoffending data, resulting in a final sample of 4,898 prisoners. The aim of the study was to examine the links between resettlement factors and reoffending during a period of two years following release from custody.

Raynor, Kynch, Roberts and Merrington (2000) used initial assessment and reconviction data of 1,115 offenders (assessed using the Assessment Case Management and Evaluation (ACE) tool) and 1,021 offenders (assessed using Level of Service Inventory – Revised (LSI-R)) to evaluate the effectiveness of the two instruments in predicting reconviction over a 12 month follow-up period.

International studies

Larney and Martire (2010) evaluated the Magistrates Early Referral Into Treatment (MERIT) programme, a voluntary diversionary programme (as part of the bail process) for people with substance use problems, based in New South Wales, Australia. The analysis included a total sample of 1,160 MERIT participants. Factors relevant to recidivism among those participants were assessed. Recidivism records were obtained for the period August 2004 to December 2007 and matched to MERIT administrative data as well as the Reoffending Database (ROD). The follow-up period ranged from 0 to 1,268 days.

Van Voorhis, Wright, Salisbury and Bauman (2010) tested a series of gender-responsive assessment models to assess their contribution to widely established gender-neutral risk needs assessments such as LSI-R. The number of female offenders varied depending on the different types of samples. The study included: (i) three prison samples (in Colorado (n=156), Minnesota (n=198) and Missouri (n=272), US), (ii) three probation samples (in Maui (n=158), Minnesota (n=233) and Missouri (n=313)) and (iii) two pre-release samples (in Colorado (n=134) and Missouri (n=162)). For the probation samples, recidivism was measured through arrest or incarceration at one or two years following assessment.

Brennan, Dieterich and Ehret (2009) examined the reliability (internal consistency) and predictive validity of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a fourth-generation risk need assessment tool, on a sample of 2,328 male and female pre-sentence investigation and probation intake cases across 18 county-level probation agencies in an eastern state (US). COMPAS assessment data were subsequently matched with computerised official criminal history records. The follow-up period ranged from 0 to 1,620 days, covering the number of days from COMPAS assessment (January 2001–December 2004) to first failure or end of the study (March 2006), whichever occurred first.

Brown, St. Amand and Zamble (2009) assessed the extent to which static and dynamic risk factors could predict criminal recidivism. They collected data on 136 adult male offenders released from Canadian federal prisons using interviews, a file review, and self-report questionnaires. A combination of pre-existing and newly developed measures was used to assess dynamic risk factors. Recidivism data (conditional release revocation) was collected for an average 10 month follow-up period.

Hsu, Caputi and Byrne (2009) examined archival LSI-R assessments from 2004 to 2007 to assess the need profiles and validity of the LSI-R for a sample of 78,052 Australian offenders serving a custodial, community or a combination of community and custodial orders. Reoffending data were obtained from the Offender Information Management System (OMIS) database from the New South Wales Department of Corrective Services.

Lemke (2009) examined the predictors of post re-release recidivism in a sample of 299 offenders (from various correctional facilities across Ohio (US)), using a structured interview guide and a self-report measure, to construct a risk/need assessment tool that correctional agencies could use to determine the placement of offenders soon to be released into the community. Offenders were tracked for an average time of 13 months after release.

Manchak, Douglas and Siranosian (2009) compared the effectiveness of the LSI-R in predicting recidivism (conviction of any new offence) for 70 female and 1,035 male offenders who had been convicted of serious violent offences, using a retrospective LSI-R database from the Washington State

(US) Institute of Public Policy. The database also contained recidivism data for all offenders who had reoffended within the 12 months follow-up period.

Salisbury, Wright and Van Voorhis (2009) administered a number of assessment tools (an Intake Custody Classification Instrument, the LSI-R and various specifically designed gender-responsive scales) to two samples of female offenders (intake sample=156 and release sample=134) in a US western state. The women were followed up at six months while in custody and for up to 44 months following release into the community. Recidivism data included prison misconducts, community recidivism and violations of supervision conditions.

Capaldi, Kim and Owen (2008) used the longitudinal Oregon Youth Study to examine 206 young men at risk of offending and reoffending in Oregon (US). They looked at the role of romantic relationships with women in men's early adulthood offending, in order to evaluate crime persistence and crime onset. They used risk data from interviews and questionnaires, and recidivism data from official County records. A follow-up period of one year was used.

Kim, Joo and McCarty (2008) examined a database with interview and survey data from a sample of 273 Day Reporting Center (DRC) participants from Douglas County, Nebraska (US) – the DRC being an intensive community-based alternative to custody. The purpose of the study was to establish the accuracy and effectiveness of the risk assessment for the DRC, in terms of termination from the programme and recidivism following release. Recidivism was recorded for all successful participants during a one year post-release follow-up period.

Lowenkamp, Lemke and Latessa (2008) outlined the construction and validation of a pre-trial assessment using multiple pre-trial agencies across two US states. A sample of 342 volunteer adult defendants completed a self-report questionnaire and a structured interview. Following data collection, the sample was divided into two separate and equal sub-samples: construction and validation. The construction sub-sample was used to identify risk factors for recidivism; the validation sub-sample was used to validate the findings.

Yahner and Visher (2008) documented the lives of 145 released offenders from Illinois (US) prisons and tracked them for three years through personal interviews and official reincarceration records. They specifically examined the factors that helped prisoners successfully avoid rearrest, reconviction and reincarceration following three years after their release from prison.

Lowenkamp and Bechtel (2007) examined the predictive validity of the LSI-R on a sample of 1,145 probationers and parolees from the State of Iowa (US) who completed LSI-R assessments between May and November 2003.

Reisig, Holtfreter and Morash (2006) explored the LSI-R's predictive accuracy for female offenders. They interviewed 235 female offenders under community supervision in Oregon and Minnesota (US) and classified them based on a pathway to crime framework. The classification included four gendered pathways (street women; drug-connected; battered; and harmed and harming) as well as two gender neutral pathways (economically motivated and unclassified). An LSI-R assessment was completed for each offender prior to their community supervision and the sample was followed up for 18 months to measure recidivism (covering violation of supervision conditions, rearrest, reconviction and revocation of community supervision).

Ostrom, Kleiman, Cheesman, Hansen and Kauder (2002) examined the relationships between individual factors of the Virginia risk assessment instrument (Worksheet D) and general recidivism among 555 prisoners in Virginia (US). Data were collected from paper case files in local probation offices, maintained by the Department of Corrections, using data collection instruments developed for the study. The follow-up period averaged 24 months.

Winters and Hayes (2001) examined the validity of the Risk Needs Inventory (RNI), an adapted version of the LSI-R, used by the Queensland Department of Corrective Services (Australia), to determine its ability to distinguish between low, medium and high risk offenders on a cohort of 600 male and female offenders on parole, probation or prison probation (i.e. a combination of prison and probation terms). The RNI was assessed against retrospective data collected from offenders' files. Reoffending was examined over a follow-up period standardised at two and four years.

Violent reoffending outcomes only

UK studies

Coid *et al*. (2011) examined the independent predictive ability of individual items within three risk assessment tools for violent recidivism. The three tools were (i) Historical-Clinical-Risk Management-20 (HCR-20), (ii) Psychopathy Checklist-Revised (PCL-R) and (iii) Violence Risk Appraisal Guide (VRAG). They also explored the possibility of improving predictive accuracy by using only the independently predictive items. A total of 1,353 prisoners in England and Wales were interviewed. The sample was followed up post-release for a mean period of almost two years. The Police National Computer (PNC) was used to obtain reoffending data.

Liu, Yang, Ramsey, Li and Coid (2011) explored violent recidivism in a prospective sample of 1,225 male prisoners in England and Wales who were serving at least two years for a sexual or violent offence. Prisoners were interviewed during a 6–12 month period prior to their release date using a series of clinical and risk assessment tools for violent and other criminal behaviour (Structured Clinical Interview for DSM-IV; Historical Clinical Risk Management-20 (HCR-20)). PNC records were used to identify violent reoffending over a four year post-release follow-up period.

Ullrich and Coid (2011) assessed the nature of the relationship between different levels of risk and protective factors with violent recidivism for male prisoners. The study consisted of two phases. In Phase I, interviews were conducted with a sample of 1,396 prisoners using PCL-R, Violence Risk Appraisal Guide (VRAG), Risk Matrix 2000/Violence Scale, HCR-20 and Offender Group Reconviction Scale 2 (OGRS2). In Phase II, a sample of 813 prisoners completed a questionnaire, devised by the lead investigator, which included 15 protective factors. PNC data was used to generate post-release reoffending measures. The mean follow-up period was 5.3 years.

International studies

Tikkanen *et al*. (2010) assessed violent recidivism in a sample of 174 male alcoholic violent offenders in Finland during an eight year non-incarceration follow-up. Offenders' psychiatric problems and alcohol consumption were identified from questionnaires and interviews, and violent recidivism data was obtained from the Legal Register Centre. The role of the enzyme monoamine oxidase genotype A (MAOA)[201] altering the effects of heavy drinking and childhood physical abuse on the risk of committing impulsive recidivistic violent crimes was also considered.

Davies and Dedel (2006) outlined the construction and validation of a Violence Risk Screening Instrument. Using a retrospective sampling frame, they identified a final sample of 385 offenders who had been reconvicted of a violent offence. The pilot instrument was completed for the 385 offenders using offender case files records and information contained in various databases. Outcome data were gathered from the Multnomah County's (US) warehouse. The data included all arrests and convictions within three years of the date of admission to community supervision.

Stalans, Yarnold, Seng, Olson and Repp (2004) examined whether the risk factors for violent recidivism were different between three groups of violent male offenders: (i) family only aggressors, (ii) non-family only aggressors and (iii) generalised aggressors (arrested for both domestic violence and other crimes), using a classification tree analysis. Probation officers recorded the case file information of 1,344 violent offenders on probation in the state of Illinois (US). Offenders' information was subsequently matched with their Illinois Criminal History Record Information to obtain criminal history and recidivism data. Violent recidivism was defined as any new arrest for a violent crime while serving their probation sentence (average length of 616 days).

Allan and Dawson (2002) identified those risk factors that typified Western Australian indigenous offenders who violently reoffend. From this analysis, they devised a population specific risk assessment tool. Using a retrospective design, they identified adult male Western Australian Indigenous offenders who required either a violent or sexual programme and randomly assigned them to three studies: (i) a predictor isolation study (n=525), (ii) a model building study (n=380) and

[201] MAOA is an enzyme that degrades amine neurotransmitters, such as dopamine, norepinephrine, and serotonin.

(iii) a cross-validation study (n=354). The follow-up period ranged from seven to nine years depending on the offence type (violent, family violence, sexual (non violent), and sexual and violence).

12.3 Results

The results are presented in line with the study's objectives and research questions, starting with the risk and protective factors for general reoffending followed by those for violent reoffending. The findings are further aligned to the following domains within the core OASys assessment, with any factors not falling within these domains then being presented separately.

- Accommodation
- Education, training and employment (ETE)
- Financial management and income
- Relationships
- Lifestyle and associates
- Drug misuse
- Alcohol misuse
- Emotional wellbeing
- Thinking and behaviour
- Attitudes

As can be seen from the following tables, some studies reported much more specific risk/protective factors than others, e.g. 'expected accommodation on release' rather than 'accommodation problems' – the detail is provided wherever possible. It is also notable that some factors are potentially more transitory than others and that the links between some factors with reoffending could be influenced by their definitions being linked to past offending (e.g. "criminal personality" and "psychopathic personality"). Information on the risk factors which were controlled for in the analysis is also provided in the tables where known. Individual studies are listed more than once in a table where associations were found for some specific risk factors but not others.

Risk factors for general reoffending

Accommodation

As set out in Table 12.1, nine studies examined whether accommodation problems were risk factors for general reoffending. A large scale UK study found that general accommodation problems (i.e. not having an address arranged on release) were associated with reoffending when combined with employment problems (i.e. not having a job arranged on release) and when controlling for a range of other risk factors. The risk of recidivism was 1.4 times higher for offenders with accommodation plus employment problems than offenders without such problems (May *et al*., 2008). However, accommodation problems on their own (e.g. expected accommodation on release) were not associated with reoffending. The findings from the other studies were inconsistent, with a large scale

US study failing to find an association between 'residential instability' and rearrest when controlling for a range of other risk factors (Brennan *et al*., 2009).

Table 12.1: Accommodation problems as risk factors for general reoffending

Study	**Sample (tool/risk score used in the study)**	**Reoffending measure**	**Other risk factors controlled for**	**Reported risk factor(s)**
Risk factors found to be significantly associated with general reoffending				
Hsu *et al*., 2009	78,052 Australian offenders (LSI-R)	Return to the care and supervision of the New South Wales Department of Corrective Services	Age, gender and sentence type	Accommodation
May *et al*., 2008	4,898 UK prisoners	Reoffending in the two years after release from prison	Demographics, criminal history, substance use or drug problem before custody, Education Training Employment (ETE) status, family ties and interventions received during custody	Accommodation (not having an address arranged on release) plus employment (not having a job arranged on release) problems
Yahner and Visher, 2008	145 US violent male offenders	Reincarceration three years after release from prison	Age, criminal history, post-prison employment, physical and mental health, family and peer relationships	Returning to live in previous neighbourhood after release
Risk factors not found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	Arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Residential instability, i.e. number of recent moves, homelessness, absence of a verifiable address
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Accommodation problems
Lowenkamp *et al*., 2008	342 US offenders (Pre-trial screening tool)	(i) Failure to attend court order (ii) New offence during release pending sentence	None	Residential instability

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor(s)
May *et al*., 2008	4,898 UK prisoners	Reoffending in the two years after release from prison	Demographics, criminal history, substance use or drug problem before custody, Education Training Employment (ETE) status, family ties and interventions received during custody	Accommodation lost Expected accommodation on release Type of accommodation before custody
Raynor *et al*., 2000	881 / 749 UK offenders (ACE / LSI-R)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, ACE / LSI-R components	Accommodation and neighbourhood
Salisbury *et al*., 2009	156/134 female US offenders (LSI-R)	Reoffence and technical violation while on parole 44 months following release	Time at risk, demographic characteristics	Accommodation problems
Winters and Hayes, 2001	600 Australian offenders (RNI)	Reoffending within two to four years	Order type, gender	Accommodation problems

Education, training and employment

As set out in Table 12.2, 13 studies examined whether education and employment, either independently or in combination, were risk factors for general reoffending. A UK study found that no continuous employment was predictive of reoffending when controlling for other static and dynamic risk factors (Raynor *et al*., 2000). As noted in the previous sub-section, in another UK study, May *et al* (2008) found that there was an interaction between employment and accommodation problems so that 'not having a job arranged on release' was associated with reoffending when combined with 'not having an address arranged on release'. They also found differences in reoffending rates by categories of employment status. Those with a paid job to go to had the lowest reoffending rate at 45% (n = 1,105) whilst those who reported not wanting to work or train had the highest reoffending rate at 75% (n = 103).

Of the large scale international studies, Hsu *et al*. (2009) found that employment/education was predictive of a return to Correctional Services supervision in a sample of nearly 80,000 Australian offenders when controlling for age, gender and sentence type. Brennan *et al*. (2009) controlled for a range of static and dynamic factors and found, in a sample of over 2,000 US offenders, that job skills, current employment and employment history were all predictive of further arrests. But the international evidence was not wholly consistent, with some of the smaller scale studies failing to find significant relationships between measures of education/employment and reoffending.

Table 12.2: Education, training and employment problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Job skills Current employment Employment history
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Employment problems
Hsu *et al*., 2009	78,052 Australian offenders (LSI-R)	Return to the care and supervision of the New South Wales Department of Corrective Services	Age, gender and sentence type	Employment/education
Lowenkamp *et al*., 2008	342 US offenders	(i) Failure to attend court order (ii) New offence during release pending sentence	None	Employment status at time of arrest
May *et al*., 2008	4,898 UK prisoners	Reoffending in the two years after release from prison	Demographics, criminal history, accommodation status, substance use or drug problem before custody, family ties and interventions received during custody	Employment (not having a job arranged on release) problems when associated with accommodation (not having an address arranged on release) Work intentions after release (paid job to go to; training or education to go to; looking after home or family; long-term sick or disabled; looking for job or course; do not want to work or train; other including retired)
Raynor *et al*., 2000	702 UK offenders (LSI-R)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, LSI items	No continuous employment
Salisbury *et al*., 2009	156/134 female US offenders (LSI-R)	Reoffence or technical violation while on parole 44 months following release	Time at risk, demographic characteristics	Education/employment

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Siddiqi, 2006	27,603 US pre-trial offenders (phase 1)	Rearrest for a violent felony offence	Age at initial arrest, ethnicity, offence type at initial arrest, prior convictions, type of release, court of disposition	Employed/school/ training
Risk factors not found to be significantly associated with general reoffending				
Lemke, 2009	299 US offenders (interview guide and self-report measure)	Any new arrests for an average of 13 months after release	Demographic and criminal history, marital status, criminal attitudes/behaviour	Level of education Ever suspended or expelled Employed at time of arrest Quit a job without having another
Ostrom *et al*., 2002	555 US offenders (Virginia risk assessment instrument).	(i) Any new felony or misdemeanour arrests (ii) Any new felony or misdemeanour convictions Follow-up period averaged 24 months	Gender, age, offence type, marital status, pilot site	Employment status
Piquero *et al*., 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics (e.g. ethnicity, employment status), heroin use, alcohol use, and exposure time	Married with full time employment
Reisig *et al*., 2006	235 US female offenders (LSI-R)	(i) Violation of supervision conditions; (ii) Rearrest; (iii) violent reconviction; or (iv) revocation of community supervision during 18 months follow-up	Age, ethnic minority, LSI-R risk need score, time at risk	Education level
Winters and Hayes 2001	600 Australian offenders (RNI)	Reoffending within one to four years	Order type, gender	Employment/education

Financial management and income

As set out in Table 12.3, six studies examined whether financial problems were risk factors for general reoffending. A UK study found that, when controlling for other risk factors, offenders with financial-related issues were 1.5 times more likely to recidivate than those with no such problems (Raynor *et al*., 2000). The study also found overlaps with other risk factors, i.e. offenders who were unemployed or had financial problems were more likely to be engaged with the benefits system or have less accommodation security, thus increasing the risk of reconviction. Two large-scale US studies (Brennan *et al*., 2009; Manchak *et al*., 2009) also found that financial problems and low wages were associated with rearrest/reconviction when controlling for a range of other factors. But the international evidence was not consistent, with three smaller scale studies failing to find an association (Brown *et al*., 2009; Salisbury *et al*., 2009; Winters and Hayes, 2001).

Table 12.3: Financial problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Financial problems Low wages
Manchak *et al*., 2009	1,105 US male and female violent offenders (LSI-R)	Conviction of any new offence in the state of Washington during a follow-up of 12 months	Gender, LSI-R scales	Financial problems
Raynor *et al*., 2000	702 UK offenders (LSI-R).	Reconviction during a 12 month follow-up period	OGRS2 scores, area, LSI-R items	Financial problems Low fixed income
Risk factors not found to be significantly associated with general reoffending				
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Financial problems
Salisbury *et al*., 2009	156/134 female US offenders (LSI-R)	Reoffence or technical violation while on parole 44 months following release	Time at risk, demographic characteristics	Financial
Winters and Hayes, 2001	600 Australian offenders (RNI)	Reoffending within one to four years	Order type, gender	Financial

Relationships

As set out in Table 12.4, 13 studies examined whether relationship and family problems were risk factors for general reoffending, utilising a range of differing measures. A large-scale UK study found that, when controlling for other static/dynamic risk factors and interventions received, not having visits from partners or family members while in prison increased the risk of general recidivism by 1.4 times (May *et al*., 2008). A large-scale US study also found that family criminality was associated with rearrest when controlling for a range of other factors (Brennan *et al*., 2009). Looking across the other studies, the findings were inconsistent. For example, one of the US studies looked at a combination of marriage and full time employment, failing to find an association with reoffending when controlling for demographic characteristics and drug/alcohol misuse (Piquero *et al.*, 2001).

Table 12.4: Relationship problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Family criminality
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Single/unsupportive partner
Capaldi *et al*., 2008	206 young male US offenders	Number of arrests in the first year after assessment	Prior arrest history, age, substance use, depressive symptoms, deviant peer associations	Female partners' anti-social behaviour
May *et al*., 2004	4,898 UK prisoners	Reoffending in the two years after release from prison	Personal characteristics, criminal history, accommodation status, substance use or drug problem before custody, ETE status, interventions received during custody	No family visits while in prison
Van Voorhis *et al*., 2010	704 female US probationers across three sites (LSI-R)	Arrests/ incarceration at one/two years following assessment	Gender-neutral risk factors, gender-responsive risk/need factors, strengths	Parental stress
Yahner and Visher, 2008	145 US violent offenders	Reincarceration three years after release from prison	Age, criminal history, post-prison employment, housing, neighbourhood characteristics, physical and mental health	Family violence or conflict before incarceration
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI).	Reconviction and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Family/marital
Risk factors not found to be significantly associated with general reoffending				
Lemke, 2009	299 US offenders (interview guide and self-report measure)	Any new arrests for an average of 13 months after release	Demographic and criminal history, education, employment, criminal attitudes/behaviour	Marital status

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Lowenkamp and Bechtel, 2007	1,145 US offenders (LSI-R)	Rearrest for felony charge or indictable misdemeanour	Sex, race, age, supervision status, time at risk and LSI-R total score	Marital status
Ostrom *et al.*, 2002	555 US offenders (Virginia risk assessment instrument)	(i) Any new felony or misdemeanour arrests (ii) Any new felony or misdemeanour convictions Follow-up period averaged 24 months	Gender, age, offence type, employment status, pilot site	Marital status
Piquero *et al.*, 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics (e.g. ethnicity, employment status), heroin use and alcohol use, exposure time	Married with full time employment
Raynor *et al.*, 2000	881 / 749 UK offenders (ACE / LSI-R)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, ACE / LSI-R components	Family / Marital
Salisbury *et al.*, 2009	156/134 female US offenders (LSI-R)	Reoffence or technical violation while on parole 44 months following release	Time at risk, demographic characteristics	Family or marital problems

Lifestyle and associates

As set out in Table 12.5, nine studies examined whether lifestyle/associate issues were risk factors for general reoffending. The majority of the studies found significant associations. For example, in a UK study (Raynor *et al*., 2000) it was found that offenders with criminal acquaintances were more likely to be reconvicted within one year, when controlling for other static and dynamic risk factors. Looking across the larger scale international studies, an Australian study (Hsu *et al*., 2009) found that leisure/recreation was a significant predictor of recidivism for male offenders on either community or custodial orders when controlling for age, gender and sentence type. In a US study of male and female offenders, Brennan *et al*., 2009 found that measures of social environment, criminal associates and leisure were all associated with rearrest when controlling for other risk factor scales.

Table 12.5: Lifestyle and associates problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Social environment Criminal associates Leisure
Capaldi *et al*., 2008	206 young male US offenders	Number of arrests in the first year after assessment	Prior arrest history, age, substance use, depressive symptoms, female partner characteristics, relationship measures	Deviant peer associations
Hsu *et al*., 2009	78,052 Australian offenders (LSI-R)	Return to the care and supervision of the New South Wales Department of Corrective Services	Age, gender and sentence type	Leisure/recreation
Raynor *et al*., 2000	903 / 702 UK offenders (ACE / LSI-R)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, ACE and LSI-R items	Associates cause risk Criminal acquaintances
Winters and Hayes 2001	600 Australian offenders (RNI)	Reoffending within one to four years	Order type, gender	Social interaction
Yahner and Visher, 2008	145 US violent male offenders	Reincarceration three years after release from prison	Age, criminal history, housing, neighbourhood characteristics, post-prison employment, physical and mental health and family	Having antisocial peers after release
Risk factors not found to be significantly associated with general reoffending				
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Leisure problems Criminal associates
Mills *et al*., 2004	144 Canadian offenders (MCAA)	Recidivism within one year of release from custody	General Statistical Information on Recidivism (GSIR) score	Criminal associates
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for any offences and any reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Companions Leisure/recreation

Drug and alcohol misuse

As set out in Table 12.6, 13 studies examined whether drugs and/or alcohol misuse was associated with reoffending. In many instances, drugs and alcohol were considered together within a single measure (whereas they are separated out as distinct domains within OASys). Although the findings across all the studies were mixed, significant associations were found in the UK studies (Farrington *et* al., 2009; May *et al*., 2008; Raynor *et al*., 2010) and two of the larger international studies (Larney and Martire, 2010; Manchak *et al*., 2009). For example, May *et al* (2008) found that 75% of offenders who had reported a drug problem prior to custody went on to reoffend within a year of release. The odds of reoffending for those reporting a drug problem before custody were 1.87 times higher than those not reporting a problem. Larney and Martire (2010) reported notable differences in reoffending rates by the principal drug used – in comparison to those participants whose principal drug use was cannabis, methamphetamine users had a 37% greater risk of recidivism, users of 'other' drugs had a 51% greater risk of recidivism, whilst heroin users had a 69% greater risk of recidivism. Farrington *et al*. (2009) looked specifically at alcohol use, finding that heavy drinking at age 18 was one of the most important risk factors in predicting persistent offending compared with adolescence-limited offending.

Table 12.6: Drug or alcohol misuse problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Substance abuse
Farrington *et al*., 2009	411 male UK offenders (including 70 persistent offenders and 53 adolescence-limited offenders)	Convictions recorded in Criminal Record Office (CRO), National Identification Service (NIS) and PNC for offenders followed up from age eight to 48	Socioeconomic factors, family factors, individual factors, behavioural factors	Heavy drinking at age 18
Larney and Martire, 2010	1,160 Australian MERIT participants	Conviction of any offence. The follow-up period ranged from 0 to 1,268 days.	Completion status, gender, age, MERIT site (Sydney metropolitan, Non-Sydney metropolitan) and prior convictions	Principal drug: Heroin or methamphetamine abuse (vs cannabis and 'other' drugs)
Lowenkamp *et al*., 2008	342 US offenders	(i) Failure to attend court order (ii) New offence during release pending sentence	None	Any drug use history Severity of problems reported from use

Study	**Sample (tool/risk score used in the study)**	**Reoffending measure**	**Other risk factors controlled for**	**Reported risk factor**
Manchak *et al*., 2009	1,035 US violent offenders (LSI-R) men only	Conviction of any new offence in the state of Washington during a follow-up of 12 months	Gender, LSI-R scales	Alcohol/drug use
May *et al*., 2008	4,898 UK prisoners	Reoffending in the two years after release from prison	Personal characteristics, criminal history, accommodation status, ETE status, family ties and interventions received during custody	Drug problem before custody
Piquero *et al*., 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics (e.g. ethnicity, marital status, employment status), exposure time	Drug use (heroin)
Raynor *et al*., 2000	903 UK offenders (ACE)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, ACE items	Drugs misuse
Risk factors not found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Substance abuse
Capaldi *et al*., 2008	206 young male US offenders	Number of arrests in the first year after assessment	Prior arrest history, age, depressive symptoms, female partner characteristics, relationship measures, deviant peer associations	Drug or alcohol abuse
Piquero *et al*., 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics (e.g. ethnicity, marital status, employment status), exposure time	Alcohol abuse
Salisbury *et al*. 2009	156/134 female US offenders (LSI-R)	Reoffence or technical violation while on parole 44 months following release	Time at risk, demographic characteristics	Drug or alcohol abuse
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for any offences and any reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Drug or alcohol problems
Winters and Hayes, 2001	600 Australian offenders (RNI)	Reoffending within one to four years	Order type, gender	Addiction problems

Emotional wellbeing

As set out in Table 12.7, eight studies examined whether issues relating to emotional wellbeing, using a wide range of measures, were risk factors for general reoffending. As can be seen, the specific risk factors covered both enduring or pervasive indicators of mental disorder and potentially more transitory emotional states (e.g. tension, depression) – an important issue when considering which dynamic risk factors to include in an assessment tool. Looking across the studies as a whole, the findings were mixed for both male and female offenders. It is also notable that the majority of studies had relatively small samples.

Table 12.7: Emotional wellbeing problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Craig *et al.*, 2004	121 male violent and sexual UK offenders (SHAPS)	General reconviction at ten years follow-up	Other SHAPS scales; RM2000-V; RM2000-S; Static-99	Tension Depression Psychopathic deviate
Hsu *et al.*, 2009	78,052 Australian offenders (LSI-R)	Return to the care and supervision of the New South Wales Department of Corrective Services	Age, gender and sentence type	Emotional/personal problems
Listwan *et al.*, 2007	277 US offenders (Jessness Inventory; Salient Factor score)	New arrest over a 10 to 12 year follow-up period	Offender group (penitentiary/prison camp), race, risk time, risk score (from the Salient Factor Score)	Neurotic personality
Wormith *et al.*, 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for any offences and any reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Interpersonal and emotional scale
Yahner and Visher, 2008	145 US violent offenders	Reincarceration three years after release from prison	Age, criminal history, post-prison employment, housing, neighbourhood characteristics, physical health, family and peer relationships	Poor mental health
Risk factors not found to be significantly associated with general reoffending				
Capaldi *et al.*, 2008	206 young male US offenders	Number of arrests in the first year after assessment	Prior arrest history, age, substance use, female partner characteristics, relationship measures, deviant peer associations	Depressive symptoms
Craig *et al.*, 2004	121 male violent and sexual UK offenders (SHAPS)	General reconviction at ten years follow-up	Other SHAPS scales; RM2000-V; RM2000-S; Static-99	Anxiety Introversion

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Raynor *et al*., 2000	903 / 702 UK offenders (ACE / LSI-R)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, ACE / LSI-R items	Emotional wellbeing
Salisbury *et al*., 2009	156/134 female US offenders (LSI-R)	Reoffence or technical violation while on parole 44 months following release	Time at risk, demographic characteristics	Emotional/personal problems Mental health Self-esteem

Thinking and behaviour

As set out in Table 12.8, six studies examined whether differing thinking and behaviour measures were risk factors for general reoffending. A UK study (Raynor *et al*., 2000) found a significant association between impulsiveness and one-year reconviction when controlling for other static and dynamic risk factors; while a large scale US study (Brennan *et al*., 2009) found that criminal personality, but not current violence, was associated with rearrest when controlling for other scales.

Table 12.8: Thinking and behaviour problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Criminal personality
Craig *et al*., 2004	121 male violent and sexual UK offenders (SHAPS)	General reconviction at ten years follow-up	Other SHAPS scales; RM2000-V; RM2000-S; Static-99	Impulsivity Aggression Lying
Listwan *et al*., 2007	277 US offenders (Jessness Inventory; Salient Factor score)	New arrest over a 10 to 12 year follow-up period	Offender group (penitentiary/prison camp), race, risk time, risk score (from the Salient Factor Score)	Aggressive personality
Raynor *et al*., 2000	903 UK offenders (ACE)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, ACE items	Impulsiveness
Risk factors not found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS)	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Current violence

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Craig *et al*., 2004	121 male violent and sexual UK offenders (SHAPS)	General reconviction at ten years follow-up	Other SHAPS scales; RM2000-V; RM2000-S; Static-99	Extroversion Hostility
Lemke, 2009	299 US offenders (interview guide and self-report measure)	Any new arrests for an average of 13 months after release	Demographic and criminal history, education, employment, marital status, criminal attitudes	Uses anger to intimidate Willing to walk away
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R, LS/CMI)	Reconviction for any offences and any reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Antisocial pattern

Attitudes

As set out in Table 12.9, seven studies examined whether attitudinal problems were risk factors for general reoffending. In the single UK study, Raynor *et al*. (2000) found that when controlling for other risk factors, offenders who agreed that reoffending was inevitable had a 1.6-fold increased risk of reconviction within 12 months. In the largest scale international study, based in the US, Brennan *et al*. (2009) found that both criminal attitudes and a history of non-compliance were associated with rearrest when controlling for other scales. Lemke (2009) included differing specific measures, finding a relationship with rearrest for 'do unto others before they do unto you' but not (i) criminal pride or (ii) belief in the possibility of overcoming the past.

Table 12.9: Attitude problems as risk factors for general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with general reoffending				
Brennan *et al*., 2009	2,328 US male and female offenders (COMPAS) univariable analysis only	An arrest for (i) any offence, (ii) a person offence and (iii) a felony offence. Follow-up period ranged from 0 to 1,620 days	COMPAS subscales	Criminal attitudes History of noncompliance
Lemke, 2009	299 US offenders (interview guide and self-report measure)	Any new arrests for an average of 13 months after release	Demographic and criminal history, education, employment, marital status, criminal behaviour	Agrees do unto others before they do unto you
Raynor *et al*., 2000	903 UK offenders (ACE)	Reconviction during a 12 month follow-up period	OGRS2 scores, Area, ACE items	Thinks reoffending inevitable

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors not found to be significantly associated with general reoffending				
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Criminal self-efficacy Poor supervision compliance Positive consequences of crime[202]
Lemke, 2009	299 US offenders (interview guide and self-report measure)	Any new arrests for an average of 13 months after release	Demographic and criminal history, education, employment, marital status, criminal behaviour	Criminal pride Agrees impossible to overcome your past
Mills *et al*., 2004	144 Canadian offenders (MCAA)	Recidivism within one year of release from custody	General Statistical Information on Recidivism (GSIR) score	Entitlement Antisocial intent
Winters and Hayes, 2001	600 Australian offenders (RNI)	Reoffending within one to four years	Order type, gender	Motivation Attitudes to supervision[203]
Wormith *et al*., 2007	61 male Canadian offenders (LS/CMI)	Reconviction for any offences and any reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Attitudes

Other risk factors

There were limited findings relating to risk factors not covered in the previous sub-sections. One Australian study (Winters and Hayes, 2001) examined the relationship between physical health and reoffending, finding a negative association, i.e. the more health problems offenders experienced, the less likely they were to reoffend. It was felt that this was a reflection of poor physical health limiting offenders' opportunities to offend. Similarly, Yahner and Visher (2008) reported in a small-scale US study that nine percent of offenders with a physical health condition, compared to 39 percent of those with no such condition, were returned to custody following release. It was noted that the former group were more likely to be largely house-bound.

[202] The 'positive consequences of crime' measure was found to be related to reoffending, but not in the direction expected – individuals who generated a greater number of positive consequences of crime were less likely to have reoffended.

[203] Attitudes to supervision was found to be inversely related to reoffending – the more non-compliant offenders were less likely to have reoffended.

Protective factors

As set out in Table 12.10, eight studies identified positive factors that were associated with lower rates of general reoffending. As can be seen, these positive factors were the inverse of some of those risk factors identified in the sub-sections above, covering stable accommodation, being employed, having a stable relationship, having positive family support, having more socially responsible attitudes towards crime and having positive self-esteem.

Table 12.10: Protective factors associated with a reduced risk of general reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk/protective factors controlled for	Reported protective factor
Protective factors found to be significantly associated with general reoffending				
Brown *et al*., 2009	136 male Canadian offenders (various pre-existing and newly developed measures)	Conditional release revocation for technical violations or new offences during a mean 10 month follow-up period	17 other time dependant risk factors	Strong social support
Capaldi *et al*., 2008	206 young male US offenders	Number of arrests in the first year after assessment	Prior arrest history, age, substance use, depressive symptoms, female partner characteristics, deviant peer associations	Stable relationship
Kim *et al*., 2008	273 US offenders (Douglas County DRC risk assessment scale)	Rearrest during one year follow-up period	Individual risk factors, case risk factors, programme need factors	Employed
Piquero *et al*., 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics(e.g. ethnicity), heroin use and alcohol use, exposure time	Full-time employment Married
Salisbury *et al*., 2009	156/134 female US offenders (LSI-R)	Reoffence or technical violation while on parole 44 months following release	Time at risk, demographic characteristics	Greater use of leisure time Self-efficacy
Van Voorhis *et al*., 2010	356 female US offenders on probation and pre-release (Gender responsive risk factors)	Serious institutional misconduct or rearrests/ technical violations during a follow-up period of 17 to 24 months	Gender-neutral risk factors, gender-responsive risk/need factors, strengths	Greater self-efficacy

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk/protective factors controlled for	Reported protective factor
Yahner and Visher, 2008	145 US violent offenders	Reincarceration three years after release from prison	Age, criminal history, physical and mental health, family and peer relationships	Having own residence within two months of leaving prison Moving to live in a 'new' neighbourhood Employed 16 months after release Positive family support Good relationship quality with family, partners and children High levels of self-esteem and control over life Positive attitude towards community integration
Winters and Hayes, 2001	600 Australian offenders (RNI)	Reoffending within one to four years	Order type, gender	Married

Risk factors for violent reoffending

Accommodation

None of the studies included within this systematic review examined whether accommodation issues were risk factors for violent reoffending.

Education, training and employment

As set out in Table 12.11, five studies examined whether issues relating to education, training and employment were risk factors for violent reoffending. There is some UK evidence supporting the view that they can be risk factors; Liu *et al*. (2011) finding that any employment problems increased the likelihood of violent recidivism. But these links were not supported by other international studies, including a large-scale US study which controlled for other risk factors (e.g. Siddiqi, 2006).

Table 12.11: Education, training and employment problems as risk factors for violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with violent reoffending				
Liu *et al*., 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction over a four year follow-up period	None specified	Employment problems
Risk factors not found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Social functioning predictors	Employment status Educational status
Piquero *et al*., 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics (e.g. ethnicity, marital status), heroin use and alcohol use, exposure time	Marriage and full-time employment
Siddiqi, 2006	4,827 US rearrested offenders (phase 1)	Rearrest for a violent felony offence	Age at initial arrest, ethnicity, offence type at initial arrest, prior convictions, type of release, court of disposition	Employed/school/ training
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for violent offences and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Employment/ education

Financial management and income

As set out in Table 12.12, one US study (Stalans *et* al., 2004) found a significant association between low income and violent reoffending when controlling for other static and dynamic risk factors.

Table 12.12: Financial problems as risk factors for violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with violent reoffending				
Stalans *et al*., 2004	1,344 US violent offenders	New arrest for a violent crime while serving a probation sentence	Demographic characteristics, criminal history, current offence characteristics, type of probation/ treatment, behaviour on probation	Annual income less than $15,001

Relationships

As set out in Table 12.13, four studies examined whether relationship issues were risk factors for violent reoffending, presenting a mixed picture. The relatively large-scale UK study (Liu *et al.*, 2011) failed to find an association with violent reoffending for relationship instability. It is important to bear in mind that this systematic review had a focus upon the risk factors for all violent reoffending rather than domestic violence.

Table 12.13: Relationship problems as risk factors for violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Social functioning predictors or childhood history predictors	Relationship instability Exposure to violence/family violence from a young age
Wormith *et al.*, 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for violent offences and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Family/marital
Risk factors not found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Social functioning predictors or childhood history predictors	Had a domestic relationship Relationship characterised by jealousy or jealous behaviour Sexual abuse in childhood Physical/emotional abuse in childhood
Liu *et al.*, 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction over a four year follow-up period	None specified	Relationship instability
Piquero *et al.*, 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics (e.g. ethnicity, employment status), heroin use and alcohol use, exposure time	Marriage and full-time employment

Lifestyle and associates

As set out in Table 12.14, six studies examined whether issues relating to lifestyle and associates were risk factors for violent reoffending. A relatively large-scale US studies looked specifically at gang membership (Stalans *et al*., 2004), finding a significant association with violent reoffending when controlling for a range of other factors.

Table 12.14: Lifestyle and associates problems as risk factors for violent reoffending

Study	**Sample (tool/risk score used in the study)**	**Reoffending measure**	**Other risk factors controlled for**	**Reported risk factor**
Risk factors found to be significantly associated with violent reoffending				
Davies and Dedel, 2006	385 male US violent offenders (Pilot instrument)	Reconviction of a violent offence (misdemeanour or felony) within three years of admission	Minor violence, severe violence, institutional violence, domestic violence, frequent violence, alcohol/drug use, history of non-compliance, special consideration	Unstable lifestyle
Stalans *et al*, 2004	1,344 US violent offenders	New arrest for a violent crime while serving a probation sentence	Demographic characteristics, criminal history, current offence characteristics, type of probation/ treatment, behaviour on probation	Gang membership
Risk factors not found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Social functioning predictors	Parasitic lifestyle Promiscuous sexual behaviour Criminal associates
Liu *et al*., 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction over a four year follow-up period	None specified	Lack of personal support
Mills *et al*., 2004	144 Canadian offenders (MCAA)	Violent recidivism within one year of release from custody	General Statistical Information in Recidivism (GSIR) score	Criminal associates
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for violent offences and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Companions Leisure/recreation

Drug or alcohol misuse

As set out in Table 12.15, seven studies examined whether drug and/or alcohol misuse were risk factors for violent reoffending. While the international evidence was not consistent, two UK studies with relatively large samples reported significant associations (Coid *et al*., 2011; Liu *et al*., 2011). A study in Finland by Tikkanen *et al*. (2010) found that the drug and/or alcohol association with violent recidivism could be influenced by genetic factors in men, i.e. those demonstrating heavy alcohol use as well as having the monoamine oxidase A gene had a significantly higher likelihood of violent reoffending.

Table 12.15: Drug or alcohol misuse problems as risk factors for violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Criminogenic need predictors	Alcohol abuse Drug use
Coid *et al*., 2011	1,271 offenders in England and Wales (HCR-20)	Violent reconviction over a mean follow-up of 1.97 years	Early maladjustment, prior supervision failure, negative attitudes, impulsivity, exposure to destabilisers, non-compliance with remediation attempts	Drug or alcohol problems
Liu *et al*., 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction over a four year follow-up period	None specified	Substance use problems
Tikkanen *et al*., 2010	174 Finnish male alcoholic violent offenders	Commission of a new violent crime (after eight years of non-incarcerated follow-up)	MAOA genotype, childhood physical abuse (CPA)	Heavy drinking
Risk factors not found to be significantly associated with violent reoffending				
Davies and Dedel, 2006	385 male US violent offenders (Pilot instrument)	Reconviction of a violent offence (misdemeanour or felony) within three years of admission	Minor violence, severe violence, institutional violence, domestic violence, frequent violence, unstable lifestyle, history of non-compliance, special consideration	Drug or alcohol problems
Piquero *et al*., 2001	524 male US offenders	Violent and non-violent offending following a seven-year post-parole period	Demographic characteristics (e.g. ethnicity, marital status, employment status), exposure time	Drug use Alcohol abuse
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for violent offences and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Drug or alcohol problems

Emotional wellbeing

As set out in Table 12.16, six studies examined whether issues relating to emotional wellbeing were risk factors for violent reoffending, covering both enduring or pervasive indicators of mental disorder and potentially more transitory emotional states (e.g. tension, depression). Two large-scale UK studies (Coid *et al*., 2011; Liu *et al*., 2011) identified psychopathy as a strong predictor. In one of these studies (Liu *et al*., 2011) personality disorder was also found to be linked to violent reoffending, but not stress or major mental illness.

Table 12.16: Emotional wellbeing as risk factors for violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with violent reoffending				
Coid *et al*., 2011	1,351 offenders in England and Wales (VRAG)	Violent reconviction over a mean follow-up of 1.97 years	Younger age at index offence, non-violent offence score, history of alcohol problems, not female victim	Psychopathic personality
Liu *et al*., 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction during four year follow-up period	None specified	Personality disorder Psychopathy
Risk factors not found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Psychological/ psychiatric/treatment history	Treated mental illness Untreated mental illness Suicide attempts, self harm of suicidal ideation
Craig *et al*., 2004	121 male violent and sexual UK offenders (SHAPS)	Violent reconviction at ten years follow-up	Other SHAPS scales; RM2000-V; RM2000-S; Static-99	Anxiety Introversion Tension Depression Psychopathic deviate
Listwan *et al*., 2007	277 US offenders (Jessness Inventory; Salient Factor score)	New arrest (violent offence) over a 10 to 12 year follow-up period	Offender group (penitentiary/prison camp), race, risk time, risk score (from the Salient Factor Score)	Neurotic personality
Liu *et al*., 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction during four year follow-up period	None specified	Stress Major mental illness
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for violent offences and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Emotional subscale

Thinking and behaviour

As set out in Table 12.17, six studies examined whether issues relating to thinking and behaviour were risk factors for violent reoffending. Across these UK and international studies, various cognitive and behavioural deficits were identified as important predictors of violent reoffending. These deficits included impulsivity (Liu *et al*, 2011; Craig *et al*, 2004), aggression (2009; Craig *et al*., 2004) and having unrealistic plans/goals (Liu *et al*, 2011; Allan and Dawson, 2002).

Table 12.17: Thinking and behaviour problems as risk factors for violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Personal/emotional/cognitive orientation predictors	Unrealistic long-term goals Irresponsibility
Coid *et al*., 2011	1,347 offenders in England and Wales (PCL-R)	Violent reconviction over a mean follow-up of 1.97 years	Poor behavioural controls, criminal versatility	Boredom/stimulation
Craig *et al*., 2004	121 male violent and sexual UK offenders (SHAPS)	Violent reconviction at ten years follow-up	Other SHAPS scales; RM2000-V; RM2000-S; Static-99	Aggression Impulsivity
Liu *et al*., 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction over a four year follow-up period	None specified	Lack of insight Plans lack feasibility Impulsivity
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for violent offences and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Antisocial behaviour
Risk factors not found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Personal/emotional/cognitive orientation predictors	Becomes bored/needs stimulation Manipulative Lack of remorse Unable to deal with strong emotions Poor coping skills Impulsivity Won't accept responsibility for actions/minimisation
Coid *et al*., 2011	1,347 offenders in England and Wales (PCL-R)	Violent reconviction over a mean follow-up of 1.97 years	Poor behavioural controls, criminal versatility	Conning/manipulation

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Craig *et al*., 2004	121 male violent and sexual UK offenders (SHAPS)	Violent reconviction at ten years follow-up	Other SHAPS scales; RM2000-V; RM2000-S; Static-99	Extroversion Hostility Lying
Listwan *et al*., 2007	277 US offenders (Jessness Inventory; Salient Factor score)	New arrest (violent offence) over a 10 to 12 year follow-up period	Offender group (penitentiary/prison camp), race, risk time, risk score (from the Salient Factor Score)	Aggressive personality

Attitudes

As set out in Table 12.18, six studies examined whether attitudinal issues were risk factors for violent reoffending. The findings were not wholly consistent, although three studies – including one with a large UK sample and one with a large US sample – found treatment non-compliance/ unresponsiveness to be a predictor of future violent reoffending (Allan and Dawson, 2002; Liu *et al*., 2011; Stalans *et al*., 2004).

Table 12.18: Attitudes problems as risk factors for violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Risk factors found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Psychological/ psychiatric/treatment history	Unresponsive to treatment (for violence)
Liu *et al*., 2011	1,225 male UK sexual or violent offenders (HCR-20)	Violent reconviction over a four year follow-up period	None specified	Negative attitudes Unresponsive to treatment
Stalans *et al*. 2004	1,344 US violent offenders	New arrest for a violent crime while serving a probation sentence	Demographic characteristics, criminal history, current offence characteristics, type of probation/ treatment, behaviour on probation	Non-compliant with treatment
Risk factors not found to be significantly associated with violent reoffending				
Allan and Dawson, 2002	254 male indigenous Australian violent offenders (new risk of violence tool)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Personal/emotional/cognitive orientation predictors	Perceives violent/ sexual offending to be acceptable behaviour

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk factors controlled for	Reported risk factor
Davies and Dedel, 2006	385 male US violent offenders (Pilot instrument)	Committing a subsequent violent offence during a follow-up period of seven to nine years	Minor violence, severe violence, institutional violence, domestic violence, frequent violence, alcohol/drug use, unstable lifestyle, special consideration	History of non-compliance
Mills *et al*., 2004	144 Canadian offenders (MCAA)	Recidivism within one year of release from custody	General Statistical Information on Recidivism (GSIR) score	Entitlement Antisocial intent
Wormith *et al*., 2007	61 male Canadian offenders (PCL-R; LS/CMI)	Reconviction for violent offences and reincarceration over a ten year follow-up	PCL-R factors; LS/CMI sections; DSM-III APD (Criterion A, C & D)	Attitudes

Other risk factors

None of the studies included in this review found any significant relationships between violent reoffending and other risk factors which were not covered in the previous sub-sections.

Protective factors

Ulrich and Coid (2011), a large-scale UK study, looked not only at risk factors for violent reoffending but also at protective factors – these being derived from recommendations from clinical experts in the field and/or reviews of the literature. Using binary logistic regression analyses, they found that five factors showed significant protective effects for violence following release, but once the four levels of RM2000-V risk (low, moderate, high and very high) were included in the model, only "spare time spent with family or friends" demonstrated a significant independent effect.

Table 12.19: Protective factors associated with a reduced risk of violent reoffending

Study	Sample (tool/risk score used in the study)	Reoffending measure	Other risk/protective factors controlled for	Reported protective factor
Protective factors found to be significantly associated with violent reoffending				
Ullrich and Coid, 2011	1,396/813 UK violent offenders (PCL-R; VRAG; HCR-20; RM 2000-V; Protective factors scale)	Violent reoffending during five year post-release follow-up period	Time-at-risk, time between release and second phase interview, RM2000-V (level of risk)	Spare time spent with family or friends

Summary of main factors identified and relationship to OASys

As shown above, the review did not identify any other risk domains that would be worthwhile additions to OASys. In terms of more specific items within the domains, not all items were consistently identified and those that were most consistently identified match closely to specific OASys questions. However, gang membership is not recorded within OASys and was found to be predictive of future violent reoffending in one relatively large US study. Thus, a question on gang associations/activities could be incorporated into the current lifestyle and associates section, although further justification from UK-based studies would strengthen this argument. In previous analysis of the textual information recorded within OASys, gangs were identified as a recurrent theme not covered in the fixed-response questions (Smith-Yau, 2009). Assessors recorded individual offenders as (i) having gang associations; (ii) showing gang mentality; and or (iii) being involved in some serious criminal gang activities. To ensure that the question is fully dynamic, focusing on current rather than previous gang involvement would be preferable.

12.4 Implications

This systematic review of the literature on dynamic (stable and acute) risk factors for general and violent reoffending identified 32 primary research studies. These were heterogeneous in terms of populations, methodology and data reporting, meaning no meta-analysis or quantitative synthesis of the data was possible. Nevertheless, various risk and protective factors were identified from the studies. While the majority of the factors closely match current OASys questions, there would appear to be potential value from including a new question within OASys on gang associations/activities. Such a question could be added to the current lifestyle and associates section. To ensure that the question is fully dynamic, focusing on current rather than previous gang involvement would be preferable. When updating the accompanying OASys guidance, a clear definition of the term ‘gang’ would be required for assessors, bearing in mind the potential for very differing interpretations and the fluid way in which the term has sometimes been used.

The studies included in this review were published between January 2000 and November 2011. Moving forward, further reviews of the literature will be conducted using the same systematic approach (and potentially engaging further with experts in the field), helping to ensure that offender assessment policy within NOMS reflects the most up-to-date knowledge about risk and protective factors and takes into account developments within the research literature. As demonstrated by the current review, there is a clear need for further studies (ideally UK-based to avoid any concerns regarding transferability) identifying: (i) positive factors which are negatively correlated with reoffending as well as those which moderate the impact of specific risk factors; and (ii) whether there are differences between the dynamic risk and protective factors according to age, gender and ethnicity. Across all studies, a more consistent and transparent approach towards measuring reoffending (and the links to reoffending) would help to identify which dynamic factors are most important. Further attention also needs to be given to which dynamic factors are truly causal, where changes over time are associated with changes in future offending behaviour when other factors are held constant.

13. Compendium conclusions

This chapter focuses upon the key implications from the totality of the research presented in the previous chapters, as well as summarising the work undertaken during 2013 to further validate and recalibrate the actuarial predictors of reoffending – OGRS4 (Chapter 8), OGP2 and OVP2 (Chapter 9), and OSP (Chapter 10). The chapter also sets out how the research recommendations are being taken forward by NOMS, including the design and implementation of the actuarial Risk of Serious Recidivism (RSR) tool and its use alongside the RoSH ratings for allocating cases to appropriate community providers. Finally, the chapter sets out potential future research, recognising that the validation of a fourth generation assessment tool such as OASys should be seen as on-going so that it reflects developments in the underlying evidence-base, the latest validation methodologies and changes in reoffending patterns, while continuing to support practitioners and current operational priorities and practices.

Good accurate assessment is recognised to be the starting point for managing offenders, helping to protect the public and reduce the risk of reoffending. To verify the robustness of an assessment tool requires continuing evaluation of its reliability and validity, and the previous chapters in this compendium cover OASys-focused studies completed between 2009 and 2012, including a systematic review of the underlying evidence-base, a survey of assessors' views and experiences, and analyses of various aspects of construct validity, internal reliability, predictive validity and dynamic validity. The research findings support the further development of OASys, identifying important potential improvements to assist practitioners in making well-informed judgments and to support the continuing use of OASys data for research and management information. Furthermore, as community rehabilitation services are opened up to a diverse range of new providers, the findings will assist with the development of assessment policy, helping to ensure that protection of the public remains paramount and that resources are used as efficiently and effectively as possible across the National Probation Service (NPS) and the Community Rehabilitation Companies (CRCs).[204] Importantly, OASys has been designated an approved tool for use by CRCs.

13.1 Implications and implementation of recommendations

OASys is aligned to the existing evidence base on 'What Works' in reducing reoffending, particularly the risk, need and responsivity (RNR) principles, and the systematic review in Chapter 12 recognises the need to continue to ensure that the assessment tool reflects the most up-to-date knowledge about risk and protective factors and takes into account developments within the research literature. To further support the risk principle, identifying which offenders should receive the available interventions,

[204] The 35 Probation Trusts in England and Wales were replaced by the NPS and 21 new CRCs on 1st June 2014. The NPS handles the caseloads of high risk offenders and has responsibility for the initial assessment and allocation of all offenders, whilst the CRCs manage low and medium risk offenders and are run by new providers appointed through a competition process.

this compendium sets out the next iterations of the OGRS and OASys predictors – OGRS4 and OGP2/OVP2 (Chapters 8 and 9). It is important for these predictors to be recalibrated every few years to take into account changes in the levels and patterns of reoffending. On this occasion, the opportunity was taken to align the predictors so that both now include predictions of (i) violent proven reoffending and (ii) general (i.e. all recordable) proven reoffending – OGRS3 had no predictor of violent recidivism, while OGP1 focused on non-violent reoffending rather than all reoffending. An 'offence-free time' element was also introduced into the predictors, recognising that an offender's probability of future proven reoffending falls with time after community sentence or discharge from custody without yet reoffending (as demonstrated in Chapter 4). The predictors thus allow a more accurate comparison of offenders at different stages of community supervision, assisting with the targeting of supervision and treatment resources.

The intention is for all of these predictors to be introduced within a future OASys change control release alongside user-friendly guidance carefully explaining the offence-free time element and its impact upon offenders' scores over the periods of supervision. A stand-alone version of OGRS4 will also be introduced and the Youth Justice Board (YJB) is currently taking steps to enable the use of OGRS4 within the assessment framework for young offenders.

Chapter 10 presents further research setting out a new predictor (OSP) for sexual offences most likely to cause serious harm – those involving direct contact with victims. It was found that this new seven item predictor has the potential to improve prediction of these contact sexual offences, and it has now been incorporated within a new Risk of Serious Recidivism (RSR) tool (see Appendix H) which is being used as part of the Case Allocation System (CAS) for allocating cases to the NPS or CRCs (National Offender Management System, 2014).[205] OSP has also been used (alongside OGRS and OVP) to help segment the NOMS caseload, supplying commissioners and providers with standardised offender profiles to which delivery and investment can be matched (National Offender Management Service, 2013).

OASys includes Risk of Serious Harm (RoSH) ratings which are determined through structured professional judgement. Within Chapter 7, it was found that there were variations in these ratings between probation trusts, and that they were less predictive of 'grave' reoffences (covering homicide, attempted murder, wounding, rape, arson, robbery and aggravated burglary) than OVP1. (The research presented in Chapter 5 also found that arson, kidnapping and racially aggravated offences were well predicted by OVP1). This research indicates that public protection could be improved by increasing the influence of actuarial scores upon RoSH ratings, informing the development of the CAS which structures practitioners' judgements by applying the new RSR tool alongside the RoSH ratings.

[205] A further publication will set out the construction and validation of RSR. It generates a summary score to indicate the likelihood of an offender committing a seriously harmful reoffence (including but not restricted to homicide, wounding with intent and contact sexual offences) within two years.

The ‘What Works’ need principle ensures that criminogenic needs are the focus of targeted interventions, rather than other needs which are not related to offending behaviour. The research in Chapter 11 proposes some changes to the eight OASys criminogenic need scales which could be implemented in a future update to the OASys IT system. These changes leave 31 scored questions across the eight scales, but all the scales (bar one – lifestyle and associates) now have four questions and a 0–8 scale. There is an impact upon the criminogenic need prevalence rates across five of the scales (although a relatively small change for two of the scales). Consequently, adjustments in the allocation of resources would be required to ensure that interventions were available to address the revised criminogenic need levels.

The ‘What Works’ principle of responsivity requires that interventions should be tailored to, amongst other things, the strengths of the offender (‘individual’ responsivity). Having examined the positive, promotive and protective factors recorded within OASys, it is recommended within Chapter 6 that consideration should be given to: (i) highlighting further the importance of identifying positive as well as risk factors during OASys assessors’ training; (ii) ensuring that the recording of positive factors is carefully monitored through existing quality assurance procedures; (iii) introducing fixed response categories to encourage more systematic recording of positive factors; and (iv) distinguishing between positive factors that need to be maintained and those that need to be developed, assisting in the identification of immediately promotive/protective factors and enabling changes in status (development vs. maintenance) to be monitored.

Further proposals for potential improvements to OASys and its associated processes are based upon the survey of prison and probation assessors (Chapter 2). It was found that there was scope to improve assessor awareness of the value and workings of the actuarial reoffending predictors – this will need to be considered when designing the communications and training documents which accompany the next iterations of the predictors. It was also thought that there was scope for increased sharing of information/good practice across establishments and trusts, and that local initiatives (such as staff briefings, discussion forums and peer review support mechanisms) could be encouraged to help continue to improve the quality of OASys assessments. Further feedback around the targeting of OASys and its layers will be considered as part of an overall review of OASys and its role in assisting the management of offenders across the diverse range of community-based providers. The survey also revealed high levels of practitioner support for the OASys RoSH ratings, which helped inform the development of the CAS.

13.2 Ongoing validation and recalibration

This compendium has so far presented studies completed between 2009 and 2012. As noted above, it is necessary to continually validate OASys, and it is especially important to validate and recalibrate the actuarial predictors of reoffending – OGRS4 (Chapter 8), OGP2 and OVP2 (Chapter 9), and OSP (Chapter 10) – as levels and patterns of reoffending may change over time.

Accordingly, further studies were undertaken during 2013. These studies followed-up the NOMS community caseload of 31 March 2010, with Police National Computer (PNC) data being extracted in December 2012 to allow a two-year proven reoffending follow-up with an adequate ‘buffer period’ for reconviction and data entry of seven months. This sample included 173,643 offenders, after excluding those who could not be followed-up for two years as they were imprisoned due to pseudoreconviction. These offenders had one-year proven reoffending rates of 30% for all offences and 15% for violent offences, and two-year rates of 41% and 24% respectively.

The OGRS4/G, OGRS4/V, OGP2 and OVP2 tools were then recalibrated. For each tool, the first stage of the process was to simply refit its parameters: the final models developed in chapters 8 and 9 were run on the March 2010 dataset, so that new parameter estimates were created for the same set of risk factors. The goodness of fit of the resulting model was then checked extensively against one and two-year reoffending outcomes, examining whether predicted and actual reoffending rates were comparable across the range of scores/statuses on each type of risk factor and offender characteristic.

As a result of these checks, two changes were made to the predictive models. First, a squared term was added to the ‘Copas rate’ element of some of the models, to improve their ability to model the reoffending patterns of those with the most intense criminal careers. Second, and more importantly in presentational terms, it was concluded that the 100-point scoring system used in OGP2 and OVP2 introduced too many problems related to the underlying mathematics of the predictive models, with the need to fit two logistic regression models causing the actual and predicted rates to deviate systematically (being too high in some regions of the predicted score and too low in other regions). The new versions of the models were therefore fitted using a conventional ordinal logistic regression. This sacrifices the exact descriptions of the relative importance of each risk factor, but these can still be summarised in user guidance on the basis of the logistic regression coefficients.

Table 13.1 summarises the relative predictive validity of the refitted models, using this 31 March 2010 dataset. Tests using the DeLong *et al*. (1988) AUC comparison method described in earlier chapters confirm that the refitted versions of OGP2 and OVP2 had significantly better relative predictive validity than all other predictors of their relevant outcomes. Refitting the models did improve their predictive validity for each predictor’s primary outcome of interest, especially for OGP2 and OVP2, but reduced the validity of the general reoffending predictors for violent recidivism. This increases the necessity of using twin predictors, rather than relying upon a general predictor to cover every outcome.

Table 13.1: Relative predictive validity of the refitted predictors: Area Under Curve (AUC) statistics for all and violent proven reoffending

Predictor	General (all) reoffending	Violent reoffending
General reoffending predictors		
OGRS3	0.771	0.716
OGP1	0.776	n/a
OGRS4/G (original)	0.783	0.737
OGRS4/G (refitted)	0.786	0.729
OGP2 (original)	0.785	0.741
OGP2 (refitted)	0.791	0.735
Violent reoffending predictors		
OVP1	n/a	0.741
OGRS4/V (original)	n/a	0.757
OGRS4/V (refitted)	n/a	0.760
OVP2 (original)	n/a	0.762
OVP2 (refitted)	n/a	0.768

The 31 March 2010 dataset was also used to test the predictive validity of OSP. Given the complex process described in Chapter 10, a full refit of the model was not feasible or desirable – the focus of this exercise was rather to check that the relatively untried OSP approach remained successful. Therefore, the parameters of the final logistic regression model were refitted, but the composition of OSP's 32-point score was not challenged.

The dataset included 12,470 offenders eligible for OSP (i.e. male offenders with current or previous statutory sexual offences and/or a current offence with a sexual element or motivation). These offenders had contact sexual reoffending rates of 0.6% at one year and 1.0% at two years. For one and two-year proven sexual reoffending, the AUCs of OSP were 0.702 and 0.713 respectively. These AUCs were superior to those found for Risk Matrix 2000 (RM2000) in other studies, and it therefore remains highly probable that OSP adds predictive value over the tools currently available.[206]

13.3 Further research

The first OASys research compendium presented studies completed between 2006 and 2009, whilst this second compendium covers the 2009–2013 research programme. The validation of a fourth generation assessment tool such as OASys should be seen as on-going, ensuring that the tool reflects developments in the underlying evidence-base, the latest validation methodologies and changes in reoffending patterns, while continuing to support practitioners and current operational priorities and practices.

[206] In reporting these refitted models, the model coefficients have not been reported. This acknowledges the intention for the predictors to be recalibrated frequently, ideally on an annual cycle, which will make the repeated publication of model coefficients overly resource-intensive. Instead, parties interested in using the predictors should contact NOMS (National.Research@noms.gsi.gov.uk) to gain information on the latest versions and discuss licensing matters as appropriate to the nature of their intended use.

Following implementation, there will be value in repeating some of the studies presented in this compendium. For example, attention could be given to the extent to which OGP2 and OVP2 scores change and whether score changes predict recidivism (as set out in Chapter 4 for OGP1/OVP1). The OGP1/OVP1 study did not consider the impact of interventions (accredited and otherwise), and a new study could attempt to take such impacts into account. The evidence of OSP's predictive validity is now strong, but it does not conclusively determine whether there is a substantive difference between OSP and RM2000 in their prediction of contact sexual reoffending. A study to address this may be possible using data held by NOMS Operational Services and Interventions Group (OSIG) on sexual offenders treated in prison. Based on preliminary data analyses undertaken on the datasets already available, it would also be possible to create predictors of indecent image and paraphilia reoffending and to test them on the OSIG database, if these were considered valuable for the risk assessment and management of sexual offenders.

Due to the limitations of the data included in the PNC extract, the OGRS4 study (Chapter 8) did not check the absolute or relative predictive validity of OGRS4 among offenders of diverse ethnic groups. This should be tested with samples drawn from NOMS and youth justice case management systems, when adequate opportunities arise (particularly as Chapter 3 identified lower levels of validity for all Black, Asian and Minority Ethnic (BME) groups in relation to non-violent reoffending and for Black and Mixed ethnicity offenders in relation to violent reoffending). An important limitation of the validation of OGRS4, OGP2 and OVP2 described above is that the 31 March 2010 samples were of NOMS cases only and did not include juveniles. The YJB's implementation of OGRS4 is therefore restricted to the version described in Chapter 8, and its recalibration will require either a separate exercise or another recalibration of the predictors which is co-ordinated to include both NOMS and YJB cases.

To further support the responsivity principle, further research could be undertaken to examine the risk and promotive effects of individual questions across the OASys sections (rather than focusing solely upon the section scales – Chapter 6), as well as the risk, promotive and protective factors for discrete offender groups, e.g. female offenders and BME offenders.

Longer term, all the predictors (including the new RSR tool) will need to be re-validated to take into account any changes in the levels and patterns of reoffending (whilst also taking into account developments with more complex predictive methods such as neural networks). Furthermore, the systematic review chapter provides a standardised methodology that could be repeated at set intervals, accounting for any developments in the wider evidence base. Within the wider research literature, there is a clear need for further studies identifying: (i) positive factors which are negatively correlated with reoffending as well as those which moderate the impact of specific risk factors; and (ii) whether there are differences between the dynamic risk and protective factors according to age, gender and ethnicity. Further attention also needs to be given within the wider literature to which dynamic factors are truly causal, where changes over time are associated with changes in future offending behaviour when other factors are held constant.

References

Andrews, D.A. and Bonta, J. (2007) *The psychology of criminal conduct (4th ed.),* Cincinnati: Anderson Publishing.

Andrews, D.A. and Bonta, J. (2010) 'Rehabilitating Criminal Justice Policy and Practice', *Psychology, Public Policy and Law*, vol. 16(1): 39–55.

Aubrey, R. and Hough, M. (1997) *Assessing Offenders' Needs: Assessment scales for the Probation Service*, Home Office Research Study 166, London: Home Office.

Aye Maung, N. and Hammond, N. (2000) *Risk of Re-offending and Needs Assessment: The User's Perspective,* Home Office Research Unit Study 211, London: Home Office

Baker, K. (2004) 'Is Asset really an Asset', In Burnett, R. and Roberts, C. (eds.) *What Works in Probation and Youth Justice*, Cullompton: Willan Publishing.

Barnett, G.D., Wakeling, H.C., and 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*,* vol. 22: 443–470.

Beasley, J.O., Hayne, A.S., Beyer, K.., Cramer, G.L., Berson, S.B., Muirhead, Y. and Warren, J.I. (2009) 'Patterns of prior offending by child abductors: A comparison of fatal and non-fatal outcomes', *International Journal of Law and Psychiatry,* vol. 32: 273–280.

Berg, N. (2005) 'Non-response bias', In Kempf-Leonard, K. (ed.) *Encyclopaedia of Social Measurement vol. 2*, London: Academic Press.

Berliner, L. and Elliott, D.M. (2002) 'Sexual abuse of children' In Myers, J.E.B (ed.) *The APSAC handbook on child maltreatment (2nd edition)*, Thousand Oaks, California: Sage.

Bewley, H. (2012) *The effectiveness of different community order requirements for offenders who received an OASys assessment*, Ministry of Justice Research Series 17/12, London: Ministry of Justice.

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

Bonta, J. and Andrews, D. A. (2007) *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation 2007-06,* Ottawa, Ontario: Public Safety Canada,

http://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/rsk-nd-rspnsvty/index-eng.aspx (accessed 30 June 2014).

Bonta, J. and Andrews, D. (2010) 'Viewing offender assessment and rehabilitation through the lens of the risk-need-responsivity model', In McNeill, F., Raynor, P., and Trotter, C. (eds.) *Offender Supervision: New directions in theory, research and practice*, Cullompton: Willan Publishing.

Bonta, J., Bogue, B., Crowley, M. and Motiuk, L. (2001) 'Implementing offender classification systems: lessons learned', In Bernfeld, G.A., Farrington, D. P., and Leschied, A.W. (eds.) *Offender Rehabilitation in Practice*, Chichester: John Wiley and Sons.

Brunton-Smith, I. and Hopkins, K. (2013) *The factors associated with proven re-offending following release* from *prison: findings from Waves 1 to 3 of SPCR: Results from the Surveying Prisoner Crime Reduction (SPCR) longitudinal cohort study of prisoners*, London: Ministry of Justice.

Burney, E. and Rose, G. (2002) *Racist offences – how is the law working? The implementation of the* legislation *on racially aggravated offences in the Crime and Disorder Act 1998*, Home Office Research Study 244, London: Home Office.

Campbell, M. A., French, S. and 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,

http://www.securitepublique.gc.ca/cnt/rsrcs/pblctns/rsk-ssmnt-tls/rsk-ssmnt-tls-eng.pdf (accessed 30 June 2014).

Chitty, C. (2004) 'The impact of corrections on reoffending: conclusions and the way forward', In Harper, G. and Chitty, C. (eds.) *The Impact of Corrections on Reoffending: A review of 'What Works',* Home Office Research Study 291, London: Home Office.

Coid, J., Hickey, N., Kahtan, N., Zhang, T. and Yang, M. (2009) 'Patients discharged from medium secure forensic psychiatry services: reconvictions and risk factors', *The British Journal of Psychiatry* 2007, vol. 190: 223–229.

Copas, J. (1992, unpublished) *Statistical analysis for a risk of reconviction predictor*, *report to the Home Office*, Coventry: University of Warwick.

Copas, J. and Marshall, P. (1998) 'The offender group reconviction scale. A statistical reconviction score for use by probation officers', *Applied Statistics,* vol. 47: 159–171.

Cornwell Management Consultants (2005) *Scoping Paper for OASys User Requirements Survey*, unpublished National Probation Service report.

Craig, L.A., Beech, A.R. and Harkins, L. (2009) 'The Predictive Accuracy of Risk Factors and Frameworks', In Beech, A.R., Craig, L.A. and Bowne, K.D. (eds.) *Assessment and Treatment of Sex Offenders: A Handbook*, Chichester: Wiley-Blackwell.

Debidin, M. (ed.) (2009) *A compendium of research and analysis on the Offender Assessment* System *(OASys) 2006 – 2009*, Ministry of Justice Research Series 16/09, London: Ministry of Justice.

DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L. (1988) 'Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach', *Biometrics,* vol. 44: 837–845.

DePanfilis, D. (2006) *Child Neglect: A Guide for Prevention, Assessment and Intervention*, Washington: Child Welfare Information Gateway,

http://www.childwelfare.gov/pubs/usermanuals/neglect/neglect.pdf (accessed 30 June 2014)

Department for Transport (2004) *Drivers convicted of dangerous or careless driving and victims: what they think of driving offences and penalties,* Road Safety Research Report 46, London: Department for Transport

Department for Transport, Local Government and the Regions (2002) *Dangerous driving and the law*, Road Safety Research Report 26*,* London: Department for Transport,

http://webarchive.nationalarchives.gov.uk/+/http://www.dft.gov.uk/pgr/roadsafety/research/rsrr/theme2/dangerousdrivinglaw26.pdf (accessed 30 June 2014).

Dillman, D., Smyth, J.D. and Christian, L.M. (2009) *Internet, Mail and Mixed-Mode Surveys: The tailored design methods*, New Jersey: John Wiley and Sons.

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

Equality and Human Rights Commission (2010) *How fair is Britain? Chapter 7. Legal security*, Manchester: Equality and Human Rights Commission,

www.equalityhumanrights.com/sites/default/files/documents/triennial_review/how_fair_is_britain_-_complete_report.pdf (accessed 30 June 2014).

Farrall, S. (2002) *Rethinking What Works with offenders: Probation, social context and desistance from crime*, Cullompton: Willan Publishing.

Farrall, S. (2004) 'Supervision, motivation and social context: what matters most when probationers desist?' In Mair, G. (ed.) *What Matters in Probation*, Cullompton: Willan Publishing.

Farrington, D.P., Loeber, R., Jolliffe, D. and Pardini, D.A. (2008) 'Promotive and Risk Processes at Different Life Stages', In Loeber, R., Farrington, D., Stouthamer-Loeber, M. and White, H.R (eds.), *Violence and Serious Theft: Development and Prediction from Childhood to Adulthood*, New York: Routledge.

Fass, T.L, Heilbrun, K., Dematteo, D. and Fretz, R. (2008) ‘The LSI-R and the COMPASS: Validation data on two risk needs tools’, *Criminal Justice and Behaviour*, vol. 35: 1095–1108.

Finklehor, D., Ormrod, R.K. and Turner, H. (2007) ‘Revictimization patterns in a national longitudinal sample of children and youth’, *Child Abuse and Neglect,* vol. 31: 479–502.

Fitzgibbon, W., Hamilton, C. and Richardson, M. (2010) ‘A risky business: An examination of Irish probation officers’ attitudes towards risk assessment’, *Probation Journal, The Journal of Community and Criminal Justice*, vol. 57(2): 163–174.

Foster, J., Newburn, T. and Souhami, A. (2005) *Assessing the impact of the Stephen Lawrence Inquiry*, Home Office Research Study 294, London: Home Office.

Gendreau, P., Little, T. and Goggin, C. (1996) ‘A meta-analysis of the predictors of adult offender recidivism: what works!’, *Criminology*, vol. 34: 575–608. Doi: 10.1111/j.1745-9125.1996.tb01220.x

Gloucestershire Probation Area (2004) *eOASys User Questionnaire Analysis – July 2004*, unpublished National Probation Service report.

Gönen, M. (2007) Analysing *receiver operating characteristic curves with SAS*, Cary, North Carolina: SAS Institute.

Griffin, H. L., Beech, A. R., Print, B., Bradshaw, H. and Quayle, J. (2008) ‘The development and initial testing of the AIM2 framework to assess risk and strengths in young people who sexually offend’, *Journal of Sexual Aggression,* vol. 14: 211–225.

Grubin, D. (2008) Validation *of Risk Matrix 2000 for use in Scotland,* Report prepared for the Risk Management Authority.

Hannah-Moffat, K., Maurutto, P. and Turnbull, S. (2009) ‘Negotiated Risk: Actuarial Illusions and Discretion in Probation’, *Canadian Journal of Law and Society*, vol. 24: 391–409.

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

Hanson, R.K. and Thornton, D. (2003) *Notes on the development of the Static-2002 (User Report No.2003-01)*, Ottawa, Ontario: Solicitor General Canada.

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

Hochstetler, A., DeLisi, M. and Pratt, T. (2010) ‘Social support and feelings of hostility among released inmates’, *Crime and Delinquency*, vol. 56(4): 588–607.

Home Office (2002) *Offender Assessment System: OASys: User Manual*, London: Home Office

Home Office (2005) *The NOMS Offender Management Model*, London: Home Office.

Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression*. New York: Wiley.

House of Commons Justice Committee (2011) *The role of the Probation Service*, London: Justice Committee,

http://www.publications.parliament.uk/pa/cm201012/cmselect/cmjust/519/51902.htm (accessed 30 June 2014).

Howard, P. (2006) *The Offender Assessment System: an evaluation of the second pilot*, Home Office Findings 278, London: Home Office.

Howard, P. (2009) 'Predictive validity of OASys: Improving prediction of violent and general offending', In Debidin, M. (ed.) *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009*, Ministry of Justice Research Series 16/09, London: Ministry of Justice.

Howard, P. (2011) *Hazards of different types of reoffending*, Ministry of Justice Research Series 3/11, London: Ministry of Justice.

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

Howard, P. and Dixon, L. (2011) 'Developing an empirical classification of violent offences for use in the prediction of recidivism in England and Wales', *Journal of Aggression, Conflict and Peace Research,* vol. 3: 141–154.

Howard, P.D. and Dixon, L. (2012) 'The construction and validation of the OASys Violence Predictor. Advancing violence risk prediction in the English and Welsh correctional services', *Criminal Justice* and *Behavior,* vol. 39: 287–307.

Howard, P.D. and Dixon, L. (2013) 'Identifying change in the likelihood of violent recidivism. Causal dynamic risk factors in the OASys Violence Predictor', *Law & Human Behavior*, vol. 37: 163–174.

Howard, P., Francis, B., Soothill, K. and Humphreys, L. (2009) *OGRS 3: The revised Offender Group Reconviction Scale*, Ministry of Justice Research Summary 7/09, London: Ministry of Justice.

Howard, P. and Moore, R. (2009) 'Measuring change in risk and need over time using OASys', In Debidin, M. (ed.) *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009.* Ministry of Justice Research Series 16/09, London: Ministry of Justice.

Jessor, R., Turbin, M.S. and Costa, F.M. (1998) 'Risk and Protection in Successful Outcomes Among Disadvantaged Adolescents', *Applied Developmental Science*, vol. 2(4): 194–208.

Jones, N.J. and Brown, S.L. (2008) ‘Special feature: positive reframing: The benefits of incorporating protective factors into risk assessment protocols’, *Crime Scene*, vol. 15(1): 22–24.

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

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

Kroner, D.G., Mills, J.F., Reitzel, L.R., Dow, E., Aufderheide, D.H. and Railey, M.G. (2007) ‘Directions for violence and sexual risk assessment in correctional psychology’, *Criminal Justice and Behavior,* vol. 34: 906–918.

Kropp, P.R., Hart, S.D., Webster, C.D. and Eaves, D. (1995) *Manual for the Spousal Assault Risk Assessment Guide (SARA): 2nd Edition*, Vancouver, British Columbia: British Columbia Institute Against Family Violence.

Laub, J.H. and Sampson, R.J. (2001) ‘Understanding deviance from crime’, In Tonry, M. (ed.) *Crime and* Justice*: An annual review of research (vol. 28),* Chicago: University of Chicago Press.

LeBel, T.P., Burnett, R., Shadd, M. and Bushway, S. (2008) ‘The ‘Chicken and Egg’ of subjective and social factors in desistance from crime’, *European Journal of Criminology*, vol. 5(2): 131–159.

Lloyd, C., Mair, G. and Hough, M. (1994) *Explaining reconviction rates: a critical analysis*, Home Office Research Study 136, London: Home Office.

López-Viets, V., Walker, D.D. and Miller, W.R. (2002) ‘What is motivation to change? A scientific analysis’, in McMurran, M. (ed.), *Motivating Offenders to Change: A Guide to Enhancing Engagement in Therapy*, Chichester: John Wiley & Sons.

Luthar, S.S. (1993) ‘Annotation: Methodological and conceptual issues in research on childhood resilience’, *Journal of Child Psychology and Psychiatry*, vol. 34: 441–453.

Mair, G., Burke, L. and Taylor, S. (2006) ‘The worst tax form you’ve ever seen? Probation officers’ views about OASys’*, Probation Journal*, vol. 53(1): 7–23.

Mann, R.E., Hanson, R.K. and 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,* vol. 22: 191–217.

Maruna, S. and LeBel, T. (2003) ‘Welcome Home? Examining the Reentry Court Concept from a Strengths-based Perspective’, *Western Criminology Review*, vol. 4(2): 91–107.

McGuire, J. (1995) *What Works: Reducing Reoffending Guidelines from Research and Practice*, Chichester: John Wiley & Sons.

McNeill, F. and Weaver, B. (2010) Changing Lives? Desistance Research and Offender Management. SCCJR Project Report, No.03/2010*,* Edinburgh: Scottish Consortium for Crime and Criminal Justice.

Merrington, S. (2004) 'Assessment tools in probation', In Burnett, R. and Roberts, C. (eds.) *What Works in Probation and Youth Justice*. Cullompton: Willan Publishing.

Miller, H. A. (2006) *IORNS Inventory of Offender Risk, Needs, and Strengths Professional Manual*, Lutz, Florida: Psychological Assessment Resources, Inc.

Ministry of Justice (2010a) *Reoffending of adults: results from the 2008 cohort*, Ministry of Justice Statistics bulletin, London: Ministry of Justice.

Ministry of Justice (2010b) *Criminal statistics, England and Wales 2009 – Appendix 5: Offence classifications*, London: Ministry of Justice.

Ministry of Justice (2010c) *Multi-Agency Public Protection Arrangements Annual Report 2009/10*, London: Ministry of Justice.

Ministry of Justice (2010d) *Offender management Caseload Statistics 2009*, London: Ministry of Justice.

Ministry of Justice (2011) *Statistics on Race and the Criminal Justice System 2010*, London: Ministry of Justice.

Ministry of Justice (2012) *Proven re-offending statistics: definitions and measurement*, London: Ministry of Justice.

Ministry of Justice (2013a) *Proven Re-offending Statistics Quarterly Bulletin July 2010 to June 2011, England and Wales,* London: Ministry of Justice.

Ministry of Justice (2013b) *Transforming Rehabilitation: A Strategy for Reform,* London: Ministry of Justice.

Ministry of Justice (2014) *Transforming Rehabilitation: A summary of evidence on reducing reoffending (second edition),* London: Ministry of Justice.

Mitchell, K.J., Finklehor, D. and Wolak, J. (2001) 'Youth internet users at risk for the most serious online sexual solicitations', *American Journal of Preventive Medicine,* vol. 32: 532–537.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) *Introduction to the theory of statistics. (3rd edition),* Singapore: McGraw-Hill.

Moore, R. (2009a) *The coverage and representativeness of OASys risk and need offender profiles: 2007 probation commencements and sentenced prisoner receptions*, Ministry of Justice Research Summary 3/09, London: Ministry of Justice.

Moore, R. (2009b) 'The internal reliability and construct validity of OASys', In Debidin, M. (ed.) *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009*, Ministry of Justice Research Series 16/09, London: Ministry of Justice.

Moore, R. (2009c) *Predicting re-offending with the OASys self-assessment questionnaire*, Ministry of Justice Research Summary 5/09, London: Ministry of Justice.

Morton, S. (2009a) 'Inter-rater reliability of OASys', In Debidin, M. (ed.) *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009*, Ministry of Justice Research Series 16/09, London: Ministry of Justice.

Morton, S. (2009b) 'Measurement of OASys rates of completion', In Debidin, M. (ed.) *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009*, Ministry of Justice Research Series 16/09, London: Ministry of Justice.

Nathan, P. and Ward, T. (2001) 'Females who sexually abuse children: Assessment and treatment issues', *Psychiatry, Psychology and Law*, vol. 8: 44–55.

National Association of Probation Officers (2011) *The Probation Directory*, Kent: Shaw & Sons Ltd.

National Offender Management Service (2008) *National rules for tiering cases and associated guidance*, Probation Circular 08/2008, London: National Offender Management Service.

National Offender Management Service (2009) *MAPPA Guidance 2009*, London: National Offender Management Service.

National Offender Management Service (2011a) *Determining Pre Sentence Reports*, Probation Instruction 05/2011, London: National Offender Management Service.

National Offender Management Service (2011b) *National Standards for the Management of* Offenders, London: National Offender Management Service.

National Offender Management Service (2013) *NOMS Evidence and Segmentation: Companion document to NOMS Commissioning Intentions from 2014*, London: National Offender Management Service.

National Offender Management Service (2014) *Case Allocation*, Probation Instruction 05/2014, London: National Offender Management Service.

OASys Data Evaluation and Analysis Team (O-DEAT) (2009) 'OASys statistics: 2008 probation and prison assessments', In Debidin, M. (ed.) *A compendium of research and analysis on the Offender Assessment System (OASys) 2006 – 2009*, Ministry of Justice Research Series 16/09, London: Ministry of Justice.

Office of the Surgeon General (2001) *Youth Violence: A Report of the Surgeon General*, Washington DC: U.S. Department of Health and Human Services.

Oppenheim, A. N. (1999) *Questionnaire Design, Interviewing and Attitude Measurement, New Edition,* London: Pinter.

Quinsey, V.L., Harris, G.T., Rice, M.E. and Cormier, C.A. (1998) *Violent offenders: Appraising and managing risk,* Washington, DC: American Psychological Association.

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

Rettinger, L.J. and Andrews, D.A. (2010) 'General risk and need, gender specificity, and the recidivism of female offenders', *Criminal Justice and Behavior,* vol. 37: 29–46.

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

Robinson, G. (2002) 'Exploring risk management in probation practice: Contemporary developments in England and Wales', *Punishment and Society*, vol. 4(1): 5–25.

Rose, G. (2000) *The criminal histories of serious traffic offenders*, Home Office Research Study 206, London: Home Office.

Sadlier, G. (2010) *Evaluation of the impact of the HM Prison Service Enhanced Thinking Skills programme on reoffending: Outcomes of the Surveying Prisoner Crime Reduction (SPCR) sample*, Ministry of Justice Research Series 19/10, London: Ministry of Justice.

Sapsford, R. (1999) *Survey Research*, London: Sage.

Schoon, I. (2006) *Risk and Resilience: Adaptations in Changing Times*, Cambridge: Cambridge University Press.

Seto, M.C., Hanson, R.K. and Babchishin, K.M. (2010) 'Contact sexual offending by men with online sexual offences', *Sexual Abuse: A Journal of Research and Treatment*, vol. 23: 124–145.

Shepherd, A. and Whiting, E. (2006) *Re-offending of adults: results from the 2003 cohort*, Home Office Statistical Bulletin 20/06, London: Home Office.

Sioui, R., Thibault, J. and Conseil, A.G. (2002) *The relevance of a cultural adaptation for aboriginals of the Reintegration Potential Reassessment Scale (RPRS)*, Research Report 109, Ottawa: Correctional Service of Canada.

Smith, P., Cullen, F. T. and Latessa, E. J. (2009) 'Can 14, 737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders.' *Criminology and Public Policy*, vol. 8: 183−208.

Smith-Yau, W. (2009) 'Exploratory research on the evidence boxes in the OASys 'relationships and lifestyle/associates' sections', In Debidin, M. (ed.) *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009*, Ministry of Justice Research Series 16/09, London: Ministry of Justice.

Soothill, K., Ackerley, E. and Francis, B. (2004) 'The criminal career of arsonists', *Medicine, Science and the Law*, vol. 44: 27–40.

Soothill, K., Fitzpatrick, C. and Francis, B. (2009) *Understanding criminal careers*, Cullompton: Willan.

Soothill, K., Francis, B., Ackerley, E. and Fligelstone, R. (2002) *Murder and Serious Sexual Assault:* What *Criminal Histories Can Reveal About Future Serious Offending*, Police Research Series 144, London: Home Office.

Soothill, K., Francis, B. and Liu, J. (2008) 'Does serious offending lead to homicide? Exploring the interrelationships and sequencing of serious crime', *British Journal of Criminology*, vol. 48: 522–537.

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

Stouthamer-Loeber, M., Loeber, R., Wei, E., Farrington, D.P. and Wikström, P.H. (2002) 'Risk and Promotive Effects in the Explanation of Persistent Serious Delinquency in Boys', *Journal of Consulting and Clinical Psychology*, vol. 70(1): 111–123.

Taylor, R. (1999) Predicting *reconvictions for sexual and violent offences using the revised offender group reconviction scale*, Home Office Research Findings 104, London: Home Office.

Thornton, D. (2007) *Scoring guide for the Risk Matrix 2000.9/SVC. February 2007 version*, http://www.birmingham.ac.uk/Documents/college-les/psych/RM2000scoringinstructions.pdf (accessed 30 June 2014).

Thornton, D. (2002) 'Constructing and Testing a Framework for Dynamic Risk Assessment', *Sexual Abuse: A Journal of Research and Treatment*, vol. 14: 137–151.

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C. and Erikson, M. (2003) 'Distinguishing and combining risks for sexual and violent recidivism', *Annals of New York Academy of Sciences,* vol. 989: 225–235.

Vogel, V., de, Ruiter, C., de, Bouman, Y. and de Vries Robbé, M. (2008) *SAPROF Structured Assessment of Protective Factors for Violence Risk: Guideline for the assessment of protective factors for violent behaviour*, 8th annual IAFMHS conference paper.

Wakeling, H.C., Howard, P.D. and Barnett, G.D. (2011) ‘Comparing the validity of the RM2000 scales and OGRS3 for predicting recidivism by internet sexual offenders’, *Sexual Abuse: A Journal of Research and Treatment*, vol. 23: 146–168.

Wakeling, H.C., Mann, R.E. and Milner, R.J. (2011) ‘Inter-rater reliability of Risk Matrix 2000/s’, *International Journal of Offender Therapy and Comparative Criminology*, vol. 55: 1324–1337.

Ward, T. and Brown, M. (2004) ‘The Good Lives Model and conceptual issues in offender rehabilitation’, *Psychology, Crime & Law*, vol. 10(3): 243–257.

Wayman, J. C. (2003) *Multiple Imputation For Missing Data: What Is It And How Can I Use It?,* Paper presented at the 2003 Annual Meeting of the American Educational Research Association,

http://www.csos.jhu.edu/contact/staff/jwayman_pub/wayman_multimp_aera2003.pdf (accessed 30 June 2014).

Webster, S., Mann, R., Carter, A., Long, J., Milner, R., O'Brien, M., Wakeling, H. and Ray, N. (2006) ‘Inter-rater reliability of dynamic risk assessment with sexual offenders’, *Psychology, Crime and Law,* vol. 12: 439–452.

Wechsler, D. (1999) *Wechsler Abbreviated Scale of Intelligence (WASI)*, San Antonio, Texas: Harcourt Assessment.

Wilson, E. and Hinks, S. (2011) *Assessing the predictive validity of the Asset youth risk assessment tool using the Juvenile Cohort Study (JCS)*, Ministry of Justice Research Series 10/11, London: Ministry of Justice.

Wolfe, D. A., Jaffe, P.G. and Crooks, C.V. (2006) *Adolescent risk behaviours: Why teens experiment and strategies to keep them safe*, New Haven, Connecticut: Yale University Press.

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

Wood, M., Cattell, J., Hales, G., Lord, C., Kenny, T. and Capes, T. (2013) *Re-offending by offenders on Community Orders: Preliminary findings from the Offender Management Community Cohort Study*, Ministry of Justice Analytical Summary, London: Ministry of Justice.

Yang, M., Liu, Y., and 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, London: Ministry of Justice.

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

Youth Justice Board (2003) *Asset: An assessment framework for young people involved in the youth justice system*, London: Youth Justice Board.

Included studies in systematic review (chapter 12)

Allan, A.A. and Dawson, D. (2002) *Developing a unique risk of violence tool for Australian indigenous offenders*. Australian Capital Territory: Criminology Research Council,

http://www.criminologyresearchcouncil.gov.au/reports/200001-06.html

Brennan, T., Dieterich, W. and Ehret, B. (2009) 'Evaluating the predictive validity of the COMPAS risk and needs assessment system', *Criminal Justice and Behavior*, vol. 36(1): 21–40.

Brown, S.L., St. Amand, M.D. and Zamble, E. (2008) 'The Dynamic prediction of criminal recidivism: a three-wave prospective study', *Law and Human Behavior*, vol. 33: 25–45.

Capaldi, D.M., Kim, H.K. and Owen, L.D. (2008) 'Romantic partners' influence on men's likelihood of arrest in early adulthood', *Criminology*, vol. 46(2): 267–299.

Coid, J.W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D. and Rogers, R. (2011) 'Most items in structured risk assessment instruments do not predict violence', *Journal of Forensic Psychiatry & Psychology*, vol. 22(1): 3–21.

Craig, L.A., Browne, K.D., Beech, A. and Stringer, I.A.N. (2004) 'Personality characteristics associated with reconviction in sexual and violent offenders', *Journal of Forensic Psychiatry Psychology*, vol. 15(3): 532–551.

Davies, G. and Dedel, K. (2006) 'Violence risk screening in community corrections', *Criminology and Public Policy*, vol. 5(4): 743–770.

Farrington, D.P., Ttofi, M.M. and Coid, J.W. (2009) 'Development of adolescence-limited, late-onset, and persistent offenders from age 8 to age 48', *Aggressive Behavior*, vol. 35(2): 150–63.

Hsu, C., Caputi, P. and Byrne, M.K. (2009) 'The Level of Service Inventory-Revised (LSI-R): A useful risk assessment measure for Australian offenders?', *Criminal Justice & Behavior*, vol. 36(7): 728–740.

Kim, D-Y., Joo, H-J. and McCarty, W.P. (2008) 'Risk assessment and classification of day reporting center clients: An actuarial approach', *Criminal Justice and Behavior*., vol.35: 792–812.

Larney, S. and Martire, K.A., (2010) ‘Factors affecting criminal recidivism among participants in the Magistrates Early Referral Into Treatment (MERIT) program in New South Wales, Australia.’, *Drug and Alcohol Review*, vol. 29(6): 684–688.

Lemke, R.J. (2009) *The construction and validation of an institutional release risk and needs assessment*. Cincinnati, USA: University of Cincinnati.

Listwan, S.J., Van Voorhis, P. and Ritchey, P.N. (2007). ‘Personality, criminal behavior and risk assessment: Implications for theory and practice’, *Criminal Justice and Behavior*, vol. 24(60): 60–75.

Liu, Y.Y., Yang, M., Ramsay, M., Li, X.S. and Coid, J. (2011) ‘A Comparison of Logistic Regression, Classification and Regression Tree, and Neural Networks Models in Predicting Violent Re-Offending.’, *Journal of Quantitative Criminology*, vol. 27: 547–573.

Lowenkamp, C.T. and Bechtel, MS. (2007) ‘The predictive validity of the LSI-R on a sample of offenders drawn from the records of the Iowa Department of Corrections data management system.’, *Federal Probation*, vol. 71(3): 1–8.

Lowenkamp, C.T., Lemke, R. and Latessa, E.J. (2008) ‘The development and validation of a pretrial screening tool.’, *Federal Probation: A Journal of Correctional Philosophy and Practice*, vol. 72(3): 1–9.

Manchak, SM., Douglas, KS. and Siranosian, M. (2009) ‘Does gender moderate the predictive utility of the Level of Service Inventory-Revised for serious violent offenders?’, *Criminal Justice and Behaviour*, vol. 36(5): 425–442.

May, C., Sharma, N. and Stewart, D. (2008) *Factors Linked to Reoffending: A One-Year Follow-Up of Prisoners Who Took Part In the Resettlement Surveys 2001, 2003 and 2004,* London: Ministry of Justice, http://nicic.gov/Library/023384

Mills, J.F., Kroner, D.G. and Hammati, T. (2004) ‘The measures of criminal attitudes and associates (MCAA). The prediction of general and violent recidivism.’, *Criminal Justice and Behaviour*, vol. 31(6): 717–732.

Ostrom, B. J., Kleiman, M., Cheesman, F., Hansen, R.M. And Kauder, N.B. (2002) *Offender risk assessment in Virginia: A three-stage evaluation: Process of sentencing reform, empirical study of diversion & recidivism benefit-cost analysis*. National Center for State Courts.

http://www.vcsc.virginia.gov/risk_off_rpt.pdf.

Piquero, A. R., Brame, R., Mazerolla, P. and Haapanen, R. (2001) *Crime in emerging adulthood: Continuity and change in criminal offending*. Rockville, Maryland: National Criminal Justice Reference Service.

Raynor, P., Kynch, J., Roberts, R. and Merrington, S. (2000) *Risk and need assessment in probation services: an evaluation*. Home Office Research Studies, London: Home Office, library.npia.police.uk/docs/hors/hors211.pdf

Reisig M. D., Holtfreter, K. and Morash, M. (2006) ‘Assessing Recidivism Risk Across Female Pathways to Crime’, *Justice Quarterly,* vol. 23(3): 384–405

Salisbury, E. J., Wright, E. M. and Van Voorhis, P. (2009) *Women on Probation: Lessons from the Gender-Responsive Classification Project*. Presented at the Annual Meeting of the American Society of Criminology, St. Louis, MO.

Siddiqi, Q. (2006) *Predicting the likelihood of pretrial rearrest: An examination of New York City defendant*. New York: New York City Criminal Justice Agency, Inc.

Stalans L.J., Yarnold P.R., Seng M., Olson D.E. and Repp M. (2004) ‘Identifying three types of violent offenders and predicting violent recidivism while on probation: a classification tree analysis*.’* Law *and Human Behavior*, vol. 28(3): 253–271.

Tikkanen, R., Ducci, F., Goldman, D., Holi, M., Lindberg, N., Tiihonen, J. and Virkkunen, M. (2010) ‘MAOA alters the effects of heavy drinking and childhood physical abuse on risk for severe impulsive acts of violence among alcoholic violent offenders’, *Alcoholism, Clinical and Experimental Research,* vol. 34(5): 853–860.

Ullrich, S. and Coid, J. (2011) ‘Protective factors for violence among released prisoners – effects over time and interactions with static risk.’, *Journal of Criminology and Clinical Psychology*, vol. 79(3): 381–390.

Van Voorhis, P., Wright EM., Salisbury E. and Bauman, A. (2010) ‘Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement.’, *Criminal Justice and Behavior*, vol. 37: 261–288.

Winters, B. R. and Hayes, H. (2001) ‘Assessing the Queensland Community Corrections RNI (Risk Needs Inventory)’, *Current Issues in Criminal Justice*, vol. 12(3): 288–305.

Wormith, S.J., Olver, M.E., Stevenson, H.E. and Girard, L. (2007) ‘The long-term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment’, *Psychological Services*, vol. 4(4): 287–305.

Yahner, J. and Visher, C. (2008) *Illinois Prisoners' Reentry Success Three Years after Release,* Washington DC: Urban Institute. Justice Policy Center, http://nicic.gov/Library/023267

Glossary

Area Under Curve (AUC)

Area Under Curve (AUC) statistics are used to test relative risk. The AUC ranges from 0 to 1 in theory, with scores above 0.7 being generally good in practice, although what can be achieved depends on the sample being studied. An AUC of 0.5 is equivalent to tossing a coin, and an AUC of 1 implies perfect foresight. The AUC statistic can be understood with a real-world analogy – it is equal to the probability that a randomly chosen reoffender will have a higher score on the predictor than a non-reoffender. For example, an AUC of 0.65 implies that when 100 pairs of offenders are checked, the reoffender will (on average) have a higher score on the predictor than the non-reoffender in 65 cases.

Absolute risk

Absolute risk equates to the probability of proven reoffending. Did offenders with a given score (which generates their predicted probability) reoffend at the same rate (actual probability) regardless of their age, gender or ethnicity? Across the whole offender group, summing across all scores, was the actual reoffending percentage similar to the mean predicted rate?

Case Allocation System (CAS)

Convicted offenders are allocated to the National Probation Service (NPS) or Community Rehabilitation Companies through the CAS. The NPS has the responsibility for running the system which involves the application of the ***Risk of Serious Recidivism (RSR) tool*** alongside ***Risk of Serious Harm (RoSH)*** ratings. It is completed either pre-sentence or within one working day of sentence where the offender is in scope for probation services.

Censoring

The termination of an individual's ***follow-up*** by an event which makes subsequent reoffending impossible or substantially less likely. Censoring events checked in this compendium are: imprisonment for a new offence (other than the offence currently of interest) or a ***pseudoreconviction***, recall to custody for breach of licence conditions, and the follow-up period continuing until there is less than one year remaining to the date of ***Police National Computer*** data extraction. (This one year period is required to ensure sufficient time for reoffences to lead to caution/conviction and subsequent PNC data entry.)

Concordance Index (C)

In fixed follow-up reoffending analysis, the concordance index (C) is identical to the ***Area Under Curve (AUC)*** but has the advantage of handling variations in the length of follow-up. It measures the probability that an offender with a worse reoffending outcome had a higher predictor score than one with a better outcome (i.e. reoffending more slowly, or not reoffending at all). It is calculated by comparing: (i) every pair of reoffenders and non-reoffenders, ensuring that the non-reoffender was at-risk for at least as long as the reoffender; and (ii) every pair of reoffenders except pairs who reoffended on the same follow-up day. The latter type of comparison differentiates C from AUC;

it checks whether the earlier reoffender had a higher risk predictor score than the later reoffender. As this is more difficult than merely predicting whether or not reoffending will occur, C scores are lower than AUC scores for the same sample. C scores for different outcomes need to be reported carefully, as the ratio of ‘easy’ yes/no comparisons to ‘hard’ earlier/later comparisons is greater for less frequent outcomes.

Copas rates

The ***Copas rate*** is a term used to describe various mathematical functions which have been developed to summarise the volume and speed of an offender’s known criminal career. It is named after Professor John Copas, principal author of the first version of ***OGRS***, and a different version has been used in each version of OGRS to date. In each version, offenders with more criminal ***previous sanctions*** have a higher ‘rate’, as do those whose criminal career spans a short number of years.

Cox proportionate hazards regression

A ***survival analysis*** technique designed to estimate the effect of a number of covariates (e.g. static and dynamic ***risk factors***: see ***OASys)*** on the time until failure (e.g. ***proven reoffending***). It allows for cases leaving the ***follow-up*** at different time points due to ***censoring***.

Criminogenic need

Criminogenic needs are those individual ***risk factors*** which contribute to or are supportive of offending and which are amenable to change. In other words, they are dynamic attributes which, when changed, are associated with changes in the likelihood of reoffending.

Follow-up

The ***follow-up*** is the period of time during which the offender is at risk of reoffending, following community sentence or discharge from custody, and for which their offending behaviour can be reliably tracked using Police National Computer (PNC) data. It can vary between individuals, and can be subdivided into fixed periods (e.g. one or three months) in order to measure ***hazards*** and survival in each period.

Hazard

The probability of ***proven reoffending*** over a short time period within the ***follow-up***. The hazard for each time period is calculated only for those offenders who have not reoffended prior to this time period and have not had their follow-up ***censored*** prior to or during this time period. For example, imagine a study of the violent reoffending of 1,000 individuals. In the first month, 100 were imprisoned for a non-violent offence and, of the remainder, 90 committed a violent reoffence. The hazard in month one is 10% (90/900). Of the remaining 810, 50 were imprisoned for a non-violent offence in the second month, so we can only study 760 in month two. In month two, 38 of this group committed a violent reoffence. The hazard in month two is 5% (38/760).

Life table

A ***survival analysis*** presentation method which traces the survival (e.g. not reoffending, or not being reassessed) of an at-risk group (e.g. offenders) over the course of their ***follow-up***.

OASys Data Evaluation and Analysis Team (O-DEAT)

O-DEAT has responsibility for: (i) the ongoing development of ***OASys***; (ii) the provision of management information; and (iii) various offender research projects. The data from all completed OASys assessments are collated centrally within database managed by the team – nearly seven million assessments had been collated by the end of March 2014.

OASys General reoffending Predictor (OGP) / OASys Violence Predictor (OVP)

OGP and OVP predict the likelihood of non-violent and violent ***proven reoffending*** respectively, by combining information on the offender's static and dynamic ***risk factors***. OGP and OVP scores are reported in raw and banded form on the OASys summary sheet. These predictors (version 1) were implemented in August 2009.

OASys Sexual reoffending Predictor (OSP)

OSP is a new predictor of sexual reoffending involving victim contact, which can be calculated from limited, quickly-assessed information. The primary ***risk factors*** are sexual offending history – including several different types of sexual offence – and age. These risk factors are used to generate a 32-point score, which is used to provide estimated one and two-year reoffending probabilities.

Offender Assessment System (OASys)

OASys is a risk assessment and management system developed and used by the prison and probation services of England and Wales. It includes analysis of static (criminal history and demographics) and dynamic (social and personal) ***risk factors***, ***Risk of Serious Harm (RoSH)*** ratings, sentence and risk management planning, a self-assessment (i.e. offender-completed) questionnaire and a summary sheet. The system was initially introduced in 2001, with the most significant changes to date taking place in August 2009. These changes streamlined the original (full) assessment while introducing a new shorter (standard) assessment and a fast review facility, allowing assessors to complete review assessments in shorter timescales where there have been no significant changes for the offender in the review period. The OASys IT system was updated in 2013 through the OASys-R project.

Offender Group Reconviction Scale (OGRS)

OGRS is a static risk assessment tool, which predicts proven reoffending using actuarial scoring rules on the basis of the offender’s age, gender, current offence type and criminal history. It can be quickly scored on the basis of a summary printout from the ***Police National Computer***. It is valid for all offenders, is in use across NOMS, and may be introduced into the juvenile correctional system in the future. Version 3 was introduced into probation practice in February 2008 and prison practice in August 2009. In NOMS practice, OGRS is used in assessment situations where ***OASys*** is not available. These include the preparation of on-the-day and five-day ***Pre-Sentence Reports***, with short-sentenced prisoners and with offenders serving non-rehabilitative court orders (e.g. Community Orders with unpaid work, curfew, attendance centre, prohibited activity and/or exclusion requirements).

Persistence (in the analysis of hazards)

In this compendium, relative hazards close to 1 are reported as persistent, as the hazard for the offence type or offender group of interest persists rather than diminishes over time.

Police National Computer

The operational IT system used by the police forces of England and Wales to record data on offenders and suspected offenders. A research copy within the Ministry of Justice (MoJ) includes sufficient data to match offenders and trace ***previous sanctions*** and ***proven reoffending***.

Positive factors

Similarly to an offender's ***risk factors***, positive factors cover both internal assets and external strengths and they can have different degrees of dynamism, with some being more susceptible to change than others.

Pre-sentence reports (PSRs)

PSRs are completed by probation practitioners in order to assist sentencers and inform the sentencing process. In addition, they form a critical part of the offender management process by providing insight to an individual's offending behaviour. The legislative framework for PSRs is contained in sections 156 to 159 of the Criminal Justice Act 2003. The Act does not prescribe the most appropriate format and timescale for the report – this decision has to be made according to the purpose of the report and the information required by the court. Oral reports should be completed on the day of sentence. Where courts have not indicated consideration of a specific sentence, or there is a clear need for a more detailed assessment than a report delivered orally can provide, a written report should be produced. The level of complexity and nature of risk and needs determines whether the written report is completed on the day or within five or 15 days. While Crown court cases may require reports taking up to 15 days, the majority of reports for Magistrates' court cases are likely to be oral or produced within five days.

Previous sanctions

The previous sanction count is the number of separate occasions on which the individual has received a conviction, caution or equivalent disposal (reprimand or final warning), prior to and including the offence(s) for which they are currently sentenced. One sanction can cover many offences.

Promotive factors

Positive factors which are negatively correlated with reoffending when controlling for risk factors amount to promotive factors.

Protective factors

The terminology of protective factors has not always been applied consistently in the literature. In Chapter 6 of this compendium they are restricted to those ***positive factors*** which moderate the impact of specific risk factors, reducing the likelihood of reoffending (thus clearly distinguishing them from ***promotive factors***). A wider definition was taken in the systematic review presented in Chapter 12 to capture both protective and promotive factors.

Proven reoffending

Committing an offence after the start of a court order (Community or Suspended Sentence Order) or release from custody, which subsequently leads to a formal caution or conviction. Noncriminal breaches (e.g. for failing to attend appointments with probation staff) are not included. In the analysis of proven reoffending, the date of the reoffence rather than the caution/conviction is of principal interest. While it is more complete and less misleading than traditional measures based on date of reconviction, it will still under-record actual offending behaviour and be affected by the activities of practitioners within the criminal justice system. For example, if the police issue few cautions and secure few convictions, the reoffending rate will be very low. For convenience within the compendium, ‘proven reoffending’ is generally referred to as simply ‘reoffending’.

Pseudoreconviction

A conviction during follow-up which relates to an offence committed before the start of ***follow-up***. This is not counted as ***proven reoffending***, but will cause ***censoring*** if it leads to imprisonment.

Relative risk

Relative risk equates to the identification of lower and higher risk offenders. Within each age, gender or ethnicity group, did those with lower predicted scores reoffend at a much lower rate than those with higher scores? It is possible for ***absolute risk*** to be estimated well but relative risk badly – for example, if ***OGP*** estimated the overall non-violent reoffending rate correctly, and the rate was consistent across diverse groups for given scores, but those with low scores reoffended at a rate not much below those with higher scores. It is also possible for relative risk to be estimated well but absolute risk badly – for example, if OGP separated non-violent reoffenders and non-reoffenders well within a given group, but systematically over- or underestimated that group’s level of non-violent reoffending. Both aspects of predictive validity must therefore be checked.

Relative hazard

A term employed in this compendium to reflect the standardisation of the hazards of different types of reoffending. This makes comparison of hazards easier, given that some types of offence are more frequent than others. For every offence, the relative hazard is 1 in the first month/quarter of follow-up, and for all later months/quarters represents the ratio of that period’s hazard to the first period’s hazard (e.g. if the sixth month’s hazard is 5% and the first month’s hazard is 10%, the sixth month’s relative hazard is 0.5 (5%/10%)).

Reliability

For a risk assessment tool to be reliable, it needs to produce consistent measurements. There are several types of reliability, e.g. inter-rater reliability and internal reliability.

Risk factors

Offender characteristics which are directly related to offending behaviour. Risk factors are used in ***OASys*** to predict ***proven reoffending***, which helps allocate treatment resources. Risk factors are often divided into static (e.g. age, gender, criminal history) and dynamic (e.g. social and economic circumstances, substance misuse, cognition, attitudes) factors. Causal dynamic risk factors are those which change over time, where such change is associated with change in future offending behaviour. Acute dynamic risk factors are those which change often, while stable dynamic risk factors change less often. Time-dependant risk factors are those which vary as a strict consequence of the passage of time, e.g. offence free time. Depending on the scoring rules for a predictor, age could be included as a static or time-dependent risk factor.

Risk Matrix 2000 (RM2000)

RM2000 is an empirically-derived actuarial risk assessment tool that uses historical information about offenders to divide them into categories that should differ substantially in their rates of reconviction for sexual or other violent offences. It was developed for use in the United Kingdom with males aged 18 and over who have been convicted of a sexual offence, at least one of which must have been committed when aged 16 or over. It has three scales: RM2000/s is a prediction scale for proven sexual offending; RM2000/v is a prediction scale for proven non-sexual, violent offending; and RM2000/c is a combination of the first two scales and predicts proven sexual or non-sexual violent offending.

Risk of Serious Harm (RoSH)

Within ***OASys***, serious harm is defined as an event which is life-threatening and/or traumatic, and from which recovery, whether physical or psychological, can be expected to be difficult or impossible. The four risk levels are defined as follows:

> Low – current evidence does not indicate likelihood of causing serious harm.
>
> Medium – there are identifiable indicators of RoSH. The offender has the potential to cause serious harm but is unlikely to do so unless there is a change in circumstances, for example, failure to take medication, loss of accommodation, relationship breakdown, drug or alcohol misuse.
>
> High – there are identifiable indicators of RoSH. The potential event could happen at any time and the impact would be serious.
>
> Very high – there is an imminent RoSH. The potential event is more likely than not to happen imminently and the impact would be serious.

Risk of Serious Recidivism (RSR) tool

The RSR tool generates a summary score to indicate the likelihood of an offender committing a seriously harmful reoffence within two years. It is based on static ***risk factors***, but can include dynamic factors when they are available. It is comprised of three sub-scores, one for contact sexual reoffending (***OSP***), one for indecent image reoffending and one for non-sexual violence (see Appendix H for further details).

Sexual offences

Sexual offences range from those causing indirect (e.g. internet-related) or direct psychological harm (e.g. voyeurism, indecent exposure) to those involving intense direct physical and psychological harm (e.g. rape). In this compendium, they are divided into contact offences against children, contact offences against adults, offences involving the making, distribution, showing and advertisement of indecent images of children, and 'paraphilia' offences.

Specialists

In this compendium, the term 'specialists' is used to refer to specialisation in the very broad offence classes covered by OGP and OVP.

Survival analysis

A family of techniques which, in the context of this compendium, generally focuses on the time until ***proven reoffending*** while allowing for ***censoring***. Rather than asking, "Will the offender reoffend within *x* days?", the question is "How likely is the offender to reoffend on each day in turn, if they have not reoffended on a previous day?". Survival analysis has the presentational advantage that shows what is happening at every stage of the follow-up. It has the statistical advantage of making more efficient use of the available data than traditional reoffending analysis, by ensuring that data on all offenders is included for as long as they can be legitimately followed up, rather than including only those who can be followed up for a fixed period. It includes techniques to measure the rate of reoffending over time (***hazards*** and ***survival functions***) and to explore risk factors associated with reoffending (***Cox proportionate hazards regression***). It is also used in this compendium to examine time to reassessment.

Survival function

The survival function for month x is the proportion of offenders who have not reoffended for the type of offence of interest by the end of month x. The method of calculation adjusts for censoring events.

Validity

For a risk assessment tool to be valid, it needs to be measuring what it is intended to measure, e.g. likelihood of reoffending. ***Reliability*** is a necessary condition for validity. However, an instrument can be reliable yet invalid, i.e. if it consistently produces inaccurate measurements.

Violent offences

In this compendium (unless otherwise indicated), offences were classed as violent within the broad classification used in ***OVP***. This encompasses offences of homicide and assault, threats and harassment, violent acquisitive offences (i.e. robbery and aggravated burglary), public order, non-arson criminal damage and weapon possession offences. Earlier research has shown that all of these have similar patterns of dynamic risk factors and tend to be committed by overlapping groups of offenders with similar risk profiles.

Appendices

Appendix A:
OASys questions included within criminogenic need scales, OGP1 and/or OVP1 (OASys sections 3 to 12)

Table A1: Scored OASys questions by section

Scored OASys question	Criminogenic need scales	OGP1	OVP1
Accommodation			
3.3: Currently of no fixed abode or in transient accommodation	✓	✓	✓
3.4: Suitability of accommodation	✓	✓	✓
3.5: Permanence of accommodation	✓	✓	✓
3.6: Suitability of location of accommodation	✓	✓	✓
Employment, training and employability			
4.2: Is the person unemployed, or will be unemployed on release	✓	✓	✓
4.3: Employment history	✓	✓	✓
4.4: Work related skills	✓	✓	✓
4.5: Attitude to employment	✓	✓	✓
4.6: School attendance			
4.7: Has problems with reading, writing and/or numeracy			
4.8: Has learning difficulties			
4.9: Any educational or formal professional/vocational qualifications			
4.10: Attitude to education/training			
Relationships			
6.1: Current relationship with close family members	✓		
6.3: Experience of childhood	✓		
6.4: Current relationship with partner			
6.6: Previous experience of close relationships	✓		
6.7 Domestic violence: perpetrator and/or victim			
6.8 Current relationship status			
6.9 Parental responsibilities			
6.10 Parental responsibilities – are these a problem?			

Scored OASys question	Criminogenic need scales	OGP1	OVP1
Lifestyle and associates			
7.2: Regular activities encourage offending	✓	✓	
7.3: Easily influenced by criminal associates	✓		
7.4: Manipulative/predatory lifestyle			
7.5: Recklessness and risk-taking behaviour	✓		
Drug misuse			
8.1: Drugs ever misused			
8.4: Current drug noted	✓	✓	
8.5: Level of use of main drug	✓	✓	
8.6: Ever injected drugs	✓	✓	
8.8: Motivation to tackle drug misuse	✓	✓	
8.9: Drug use and obtaining drugs a major activity/ occupation	✓	✓	
Alcohol misuse			
9.1: Is current use a problem	✓		✓
9.2: Binge drinking or excessive use of alcohol in last 6 months	✓		✓
9.3: Frequency and level of alcohol misuse in the past	✓		
9.4: Violent behaviour related to alcohol use at any time			
9.5: Motivation to tackle alcohol misuse	✓		
Thinking and behaviour			
11.1: Level of interpersonal skills			
11.2: Impulsivity			
11.3: Aggressive/controlling behaviour			
11.4: Temper control			✓
11.5: Ability to recognise problems	✓	✓	
11.6: Problem solving skills	✓	✓	
11.7: Awareness of consequences	✓	✓	
11.8 Achieves goals			
11.9: Understands other people's views	✓	✓	
11.10: Concrete/abstract thinking			

Scored OASys question	Criminogenic need scales	OGP1	OVP1
Attitudes			
12.1: Pro-criminal attitudes	✓	✓	✓
12.3: Attitude towards staff			
12.4: Attitude towards supervision/licence	✓	✓	✓
12.5: Attitude to community/society	✓	✓	✓
12.6: Does the offender understand their motivation for offending			
12.8: Motivation	✓	✓	✓

Appendix B:
OGP1 and OVP1 dynamic risk factor scoring

Table B1 shows the questions used to score dynamic risk factors in the full and standard versions of OGP1 and OVP1, and the raw and weighted scores deriving from these. Individual questions are scored 0/2 or 0/1/2, where 2 represents a significant problem, unless otherwise stated in the table's note.

Weighted scores corresponding to intermediate raw scores can be calculated on a pro rata basis, with rounding as necessary. For example, in full OASys, the maximum raw score for attitudes is 8, and the maximum OGP1 weighted score is 5. A raw score of 3 therefore produces an OGP1 weighted score of 2 (as 3*(5/8)=1.875, which rounds to 2).

Static risk factors are scored as in Howard (2009). In full and standard OASys, two new questions – 1.24 (Cautions/reprimands/final warnings) and 1.26 (Violent convictions/cautions/ reprimands/final warnings) –allow complete counts of previous sanctions (OGRS3/OGP1) and previous violent and non-violent sanctions (OVP1) to be calculated. OGRS3 scores are automatically calculated in full and standard OASys, and reported as item 1.27.

Table B1: Dynamic risk factors scored in OGP1 and OVP1

Section	Questions	Maximum raw score	Maximum weighted score	
			OGP1	OVP1
Analysis of offences	2.6* Does the offender understand impact and consequences of offending on victim, community, wider society?	See Note	--	4
Accommodation	3.3* Currently no fixed abode or transient 3.4 Suitability of accommodation 3.5 Permanence of accommodation 3.6 Suitability of location of accommodation	8	5	4
Education, training and employability	4.2* Unemployed? 4.3 Employment history 4.4 Work-related skills 4.5 Attitude towards employment	8	5	6
Lifestyle & associates	7.2* Regular activities encourage offending	2	5	--
Drug misuse	8.4* Current drug misuse 8.5 Level of use of main drug 8.6 Injecting drugs 8.8* Motivation to tackle drug misuse 8.9 Drug use / obtaining drugs a major activity/occupation?	10	15	--
Alcohol misuse	9.1* Current [chronic] use a problem? 9.2* Binge drinking in last six months	4	--	10
Emotional wellbeing	10.7* Current/pending psychiatric treatment	See Note	--	4

Section	Questions	Maximum raw score	Maximum weighted score	
			OGP1	OVP1
Thinking & behaviour	11.5 Problem recognition 11.6* Problem solving 11.7* Awareness of consequences 11.9* Understands others views	8	5	--
Temper control	11.4* Temper control	2	--	6
Attitudes	12.1* Pro-criminal attitudes 12.4 Attitude towards supervision/licence 12.5 Attitude towards community/society 12.8 Motivation to address offending	8	5	6

Note. All questions are scored in full OASys. Questions marked * are also scored in standard OASys. The standard version uses a different scoring system, so that the maximum weight is still available for all risk factors. Question 2.6 is binary: the answer ‘does not understand impact...’ raises risk. Question 10.7 is binary: requiring treatment raises risk. While item 8.5 is not displayed to assessors in standard OASys, the information required to score it is still available in the grid of drugs used.

Appendix C:
Distribution of OGP1 and OVP1 100-point scores and one- and two-year reoffending rates

Table C1: Likelihood of proven non-violent reoffending by OGP1 score

Offender's weighted score (/100)	1 year non-violent reoffending rate (sample mean: 31%)	2 year non-violent reoffending rate (sample mean: 42%)	% of initial assessments (N=220,793)
0 to 13	Below 6%	Below 10%	9%
14 to 26	6–11%	10–19%	17%
27 to 35	12–18%	20–29%	13%
36 to 43	19–26%	30–39%	12%
44 to 49	27–34%	40–49%	9%
50 to 56	35–45%	50–59%	11%
57 to 63	46–55%	60–69%	10%
64 to 71	56–67%	70–79%	10%
72 to 84	68–82%	80–89%	8%
85 to 100	83% and over	90% and over	0.8%

Table C2: Likelihood of proven violent reoffending by OVP1 score

Offender's weighted score (/100)	1 year violent reoffending rate (sample mean: 20%)	2 year violent reoffending rate (sample mean: 30%)	% of initial assessments (N=220,997)
0 to 22	Up to 5%	Below 10%	12%
23 to 34	5 to 11%	10 to 19%	24%
35 to 41	12 to 17%	20 to 29%	20%
42 to 47	18 to 26%	30 to 39%	16%
48 to 53	27 to 33%	40 to 49%	13%
54 to 59	34 to 43%	50 to 59%	8%
60 to 65	44 to 54%	60 to 69%	4.5%
66 to 72	55 to 66%	70 to 79%	2.2%
73 to 100	67% and over	80% and over	0.7%

Appendix D: Cox regression models examining impact of changes in dynamic risk factors

Tables D1 to D3 set out the results of pairs of Cox regression models using the weighted subscales in the OGP1 or OVP1 scores. In each table, one model used only the initial scores, while the second used both initial scores and changes in scores. A 'neutral' result would show similar Beta (effect sizes) for all items in the model; finding greater Beta for some risk factors than others would suggest that the weighting of risk factors in the predictor could be revised in order to maximise predictive validity.[207] In the models including both initial score and change in score, differences in Beta between the two versions of the same dynamic risk factor would indicate that the initial score was more/less predictive than the change in score.

The initial-score-only portion of D1 shows the disadvantage of using OGP1 – designed to predict non-violent recidivism – to predict all recidivism. The risk factors which are in both OGP1 and OVP1 (accommodation, employability and attitudes) and therefore should be associated with all rather than some types of recidivism contributed more to prediction than those which are only in OGP (drug misuse and thinking/behaviour). The static score (i.e. OGRS3) contributed more to prediction than the dynamic risk factors. This is unsurprising – Howard (2009) explained that the OGP1 score was distorted to give a higher weight to dynamic risk factors than did the underlying logistic regression model, though the effect of the distortion was monitored to ensure that this caused only minimal loss of predictive validity. The time-dependent portion of Table D1 gives higher weight to the initial dynamic scores. This result, which is common across the time-dependent parts of Tables D1 to D3, may be due to the initial scores applying on average more to the earlier days of follow-up when their relevance is greater (i.e. the changes in score share some of the model variance in the later days). Changes in some dynamic items had similar per-point predictive validity to the initial scores, but changes in employability were little associated with recidivism while changes in drug misuse and most especially attitudes were more strongly associated with recidivism than the initial scores for these factors.

Table D2 features predictors of non-violent reoffending. Compared with Table D1, the drug misuse score was a far stronger predictor, but thinking and behaviour was now significantly negatively associated with reoffending (i.e. worse (higher) scores were associated with lower risk of non-violent reoffending, after taking the other risk factors into account). Changes in thinking and behaviour scores were however positively associated with changes in recidivism risk; as in Table D1, changes in employability were only weakly associated with recidivism, and changes in attitudes were very strongly associated with recidivism.

[207] It is possible that the relative importance of risk factors has changed since the creation of OGP1 and OVP1, and more likely that the comparatively small sample used to create OGP1 and OVP1 allowed some random modelling error.

In Table D3, initial psychiatric treatment was weakly associated with violent recidivism, and initial recognition of the impact of offending was very slightly negatively associated (controlling, as always, for the other scored risk factors). The static score was highly predictive – the pro-dynamic distortion in scoring cited in Howard (2009) and mentioned above was greater for OVP1 than OGP1 (although still not enough to damage predictive validity in a clinically significant manner), so the static score would be expected to have a higher weighting whenever a regression model is run and its coefficient is determined naturally. With the exception of employability, changes in score were equally or more predictive than initial scores in the time-dependent model.

Table D1: Cox regression models: risk factors in OGP1 as predictors of any reoffending

Item type	Risk factor (maximum points)	Model using initial assessment only				Model with time-dependent covariates			
		Beta	SE(Beta)	Hazard ratio		Beta	SE(Beta)	Hazard ratio	
				Point	Range			Point	Range
Static	OGRS3 score (60)	0.047	0.000	1.049	17.3	0.047	0.000	1.048	16.9
Dynamic: initial assessment	Accommodation (5)	0.031	0.002	*1.032*	1.17	0.037	0.002	1.038	1.21
	Employability (5)	0.046	0.002	1.047	1.26	0.044	0.002	1.045	1.24
	Activities (5)	0.037	0.002	1.038	1.20	0.041	0.002	1.041	1.23
	Drug misuse (15)	0.017	0.001	1.017	1.28	0.018	0.001	1.018	1.31
	Thinking & behaviour (5)	0.014	0.003	1.014	1.07	0.017	0.003	1.017	1.09
	Attitudes (5)	0.034	0.003	1.035	1.19	0.040	0.003	1.040	1.22
Dynamic: change from initial to most recent	Accommodation (5)					0.045	0.003	1.046	1.25
	Employability (5)					0.007	0.007	*1.007*	1.03
	Activities (5)					0.030	0.004	1.030	1.16
	Drug misuse (15)					0.024	0.003	1.024	1.44
	Thinking & behaviour (5)					0.043	0.006	1.044	1.24
	Attitudes (5)					0.090	0.007	1.094	1.57

Note. Only scores of 0, 3 and 5 are possible for Activities. Scores of 1, 4, 7, 10 and 13 are not possible for Drug misuse.

Table D2: Cox regression models: risk factors in OGP1 as predictors of non-violent reoffending

Item type	Risk factor (maximum points)	Model using initial assessment only				Model with time-dependent covariates			
		Beta	SE(Beta)	Hazard ratio		Beta	SE(Beta)	Hazard ratio	
				Point	Range			Point	Range
Static	OGRS3 score (60)	0.054	0.000	1.055	24.9	0.053	0.000	1.055	24.2
Dynamic: initial assessment	Accommodation (5)	0.027	0.002	1.027	1.14	0.032	0.002	1.032	1.17
	Employability (5)	0.051	0.002	1.053	1.29	0.049	0.002	1.051	1.28
	Activities (5)	0.027	0.002	1.028	1.15	0.031	0.002	1.032	1.17
	Drug misuse (15)	0.048	0.001	1.050	2.07	0.051	0.001	1.052	2.14
	Thinking & behaviour (5)	-0.019	0.003	0.981	0.91	-0.017	0.003	0.983	0.92
	Attitudes (5)	0.036	0.003	1.037	1.20	0.042	0.003	1.043	1.24
Dynamic: change from initial to most recent	Accommodation (5)					0.037	0.003	1.038	1.21
	Employability (5)					0.016	0.007	*1.017*	1.09
	Activities (5)					0.030	0.004	1.031	1.16
	Drug misuse (15)					0.039	0.003	1.039	1.79
	Thinking & behaviour (5)					0.032	0.006	1.032	1.17
	Attitudes (5)					0.097	0.007	1.102	1.63

Note. Only scores of 0, 3 and 5 are possible for activities. Scores of 1, 4, 7, 10 and 13 are not possible for drug misuse.

Table D3: Cox regression models: risk factors in OVP1 as predictors of violent reoffending

Item type	Risk factor (maximum points)	Model using initial assessment only				Model with time-dependent covariates			
		Beta	SE(Beta)	Hazard ratio		Beta	SE(Beta)	Hazard ratio	
				Point	Range			Point	Range
Static	Total static score (60)	0.076	0.001	1.079	96.2	0.075	0.001	1.078	91.9
Dynamic: initial assessment	Recognises impact (4)	-0.004	0.003	*0.996*	0.98	-0.006	0.003	0.994	0.98
	Accommodation (4)	0.041	0.003	1.042	1.18	0.051	0.003	1.052	1.23
	Employability (6)	0.058	0.002	1.059	1.41	0.057	0.002	1.058	1.40
	Alcohol misuse (10)	0.045	0.001	1.046	1.58	0.050	0.001	1.051	1.65
	Psychiatric treatment (4)	0.022	0.004	1.022	1.09	0.021	0.004	1.022	1.09
	Temper control (6)	0.034	0.002	1.034	1.23	0.040	0.002	1.041	1.27
	Attitudes (6)	0.044	0.004	1.045	1.30	0.046	0.004	1.047	1.32
Dynamic: change from initial to most recent assessment	Recognises impact (4)					0.001	0.010	*1.001*	1.00
	Accommodation (4)					0.065	0.005	1.067	1.30
	Employability (6)					0.029	0.007	1.030	1.19
	Alcohol misuse (10)					0.046	0.003	1.047	1.58
	Psychiatric treatment (4)					0.048	0.013	1.049	1.21
	Temper control (6)					0.060	0.005	1.062	1.43
	Attitudes (6)					0.062	0.008	1.064	1.45

Note. Single-point hazards are included for impact and psychiatric to compare weightings; in reality these are based on single binary questions (i.e. can only be scored 0 or 4), so only the 'Range' hazard ratio is relevant. Alcohol misuse can only be scored 0, 3, 5, 8 or 10. Temper control can only be scored 0, 3 or 6.

Appendix E:
Goodness-of-fit of static risk, dynamic risk, promotive and protective factors model

To assess the goodness-of-fit of the static risk, dynamic risk, promotive and protective factors model across risk levels, the validation sample was divided into ten equal-sized deciles based upon their predicted probabilities of reoffending. Actual and predicted proven reoffending rates were then compared to assess whether any general over- or under-estimations had occurred. As shown by Table E1, all residuals (the differences between the actual and predicted rates) were within two percentage points. However, the differences across risk levels were statistically significant (chi-square value of 24.133; p=0.004).

Table E1: Residual values from static risk, dynamic risk, promotive and protective factors model across risk levels

Grouping	n	Reoffending rate		Residual (actual minus predicted rate)
		Actual	Predicted	
1	3648	9.9%	11.7%	-1.8%
2	3646	18.8%	17.8%	1.1%
3	3647	28.1%	26.1%	2.0%
4	3652	36.9%	35.9%	1.1%
5	3647	47.0%	46.1%	0.9%
6	3650	53.8%	55.7%	-1.8%
7	3649	63.9%	63.8%	0.0%
8	3648	69.3%	70.7%	-1.4%
9	3649	78.8%	77.4%	1.4%
10	3648	86.1%	84.5%	1.6%

The reoffending rates and residuals may not match due to rounding.

Model fit was also tested for different offender groups within the validation sample, e.g. low vs. high risk of serious harm offenders. As shown by Table E2, the predicted reoffending rate was significantly different from the actual rate for a number of the offender groups. The greatest under-prediction was for Black offenders with a residual of 6.1% and the greatest over-prediction was for offenders whose current offence related to drugs with an over-prediction of 6.4%. The goodness-of-fit of the static risk, dynamic risk, promotive and protective factors model was therefore less strong for specific offender groups.

Table E2: Residual values from static risk, dynamic risk, promotive and protective factors model for various offenders groups

Grouping variable	Value	n	Reoffending rate		Residual (actual minus predicted rate)
			Actual	Predicted	
Age	18–20	6555	61.7%	59.0%	2.6%*
	21–24	7242	56.2%	54.9%	1.4%
	25–30	17000	48.9%	49.1%	-0.2%
	41+	5672	27.1%	29.4%	-2.3%*
Gender	Male	31573	50.4%	49.5%	0.9%*
	Female	4911	42.0%	45.4%	-3.4%*
Ethnicity	White	29717	50.6%	50.3%	0.3%
	Black	1131	51.9%	45.8%	6.1%*
	Asian	967	42.5%	39.5%	3.0%
	Mixed	532	54.5%	52.6%	1.9%
	Other	229	38.4%	37.2%	1.2%
Offence	Violence against the person	8726	44.2%	43.9%	0.3%
	Sexual offence	942	20.2%	24.9%	-4.8%*
	Burglary	2809	70.3%	65.7%	4.7%*
	Robbery	928	54.5%	55.4%	-0.8%
	Theft and handling	5513	66.2%	64.9%	1.2%
	Fraud and forgery	1198	33.1%	37.6%	-4.4%*
	Criminal damage	1097	58.2%	56.2%	2.0%
	Drug offences	2660	39.8%	46.3%	-6.4%*
	Other indictable offences	2837	42.4%	44.5%	-2.1%
	Summary motoring offences	7177	45.0%	42.6%	2.5%*
	Other summary offences	1922	44.5%	43.9%	0.6%
Sentence	CJA 03 Community Sentence	3617	54.1%	50.2%	3.9%
	CPO	5724	34.9%	35.3%	-0.4%
	CRO	8500	50.7%	49.1%	1.6%*
	CPRO	2418	51.4%	49.1%	2.3%
	Custody/YOI	9332	52.6%	55.3%	-2.7%*
	CJA 03 Suspended Sentence	525	49.5%	46.2%	3.4%
	Other	1629	62.1%	59.8%	2.3%
Risk of serious harm (highest community level)	Low	21210	46.8%	46.8%	0.0%
	Medium	12889	53.6%	52.2%	1.4%*
	High/very high	2385	48.4%	51.1%	-2.7%

Asterisks indicate whether rates differ significantly (confidence level *<.001). The reoffending rates and residuals may not match due to rounding.

Appendix F:
High/very high risk of serious harm checklist

In August 2009, a Pre-Sentence Report (PSR) decision tool was shared with probation areas. It was produced in order to help probation staff make an increased proportion of magistrates' court PSRs within five days or on an oral report basis (National Offender Management Service, 2011a), without giving courts inadequate information on the most serious and complex cases. One of the key principles was that five day reports (rather than 15 day reports) should still be used when the offender was more likely to be rated high or very high risk. The high/very high RoSH checklist was based on OASys assessments between 2005 and 2008, and was designed to be applied very rapidly using limited information. The information which was considered usable was mostly restricted to Sections 1 and 2 (offending information/offence analysis) of OASys. Those who scored highly on the checklist should be more likely to be rated high/very high risk, with logistic regression modelling looking at the relationship between checklist scores and whether the offender was actually rated high/very high risk.

The final version of the checklist included 10 items and was scored on a unit basis, giving each offender a score from 0 to 10 (see Table F1 below). A number of variations of the checklist were tested. Firstly, age and gender were added to the logistic regression model. Secondly, a number of extra items were tested, which would have extended the checklist beyond 10 items. Thirdly, non-unit item scoring was tested. None of these variations were considered to add a sufficient degree of predictive value to the checklist that this extra value which outweighed the loss of simplicity for practitioners.

Table F1: High/very high risk of serious harm checklist questions

Behaviour/characteristic	**Score 1 if present, score 0 if not present**
Any proven offence of murder, manslaughter, attempted murder, GBH, wounding, robbery and/or abduction?	
Current offence involves a victim aged under 16 (at time of offence)	
Any proven sexual offence, or offence with a sexual element or motivation	
Any proven arson offence, or offence involving arson	
Perpetrator of domestic abuse at any time	

Behaviour/characteristic	**Score 1 if present, score 0 if not present**
Carrying/use of a weapon in any proven offence	
Excessive/sadistic violence in any proven offence	
Does not recognise the impact and consequences of the current offence on victim, community, wider society	
Any elements of the current offence linked to risk of serious harm, risks to the individual and other risks. *NB This may include elements which have already been scored above i.e. you may ‘double count’ as appropriate.*	
Currently of no fixed abode or in transient accommodation	
Total score	

Appendix G: Modified high/very high risk of serious harm prediction model

Table G1: Logistic regression model of high / very high RoSH status based on 10-point checklist, age and gender

Parameter	Estimate	Standard error	Odds ratio
Constant	*-4.944*	*0.047*	*n/a*
Checklist score (per point)	0.989	0.004	2.69
Gender female	-0.325	0.013	0.72
Age	-0.0123	0.0025	n/a
Age squared	0.000236	0.000032	n/a

Note. All parameters were significant at p<.001. N = 516,461, number of high/very high risk cases = 44,489.

Appendix H:
The Risk of Serious Recidivism (RSR) tool

The RSR tool generates a summary score to indicate the likelihood of an offender committing a seriously harmful reoffence within two years from either the start of the community sentence or discharge from prison. It is based on static risk factors, but can include dynamic factors when they are available.

What are seriously harmful (re)offences?

Seriously harmful offences are those statutory offences which usually fall within the existing qualitative description of serious harm used by NOMS: “an event, which is life-threatening and/or traumatic, from which recovery, whether physical or psychological, can be expected to be difficult or impossible”. Through consultation and comparison with similar existing offence lists (Serious Further Offences, offences qualifying for Multi-Agency Public Protection Arrangements (MAPPA), and Ministry of Justice reoffending statistics), it was agreed that an offence was serious if any of the ten criteria below were satisfied

1. The victim died, or the offence was attempted murder or conspiracy to murder.
2. The victim suffered physical harm sufficient for a conviction of wounding or grievous bodily harm (GBH).
3. The victim was taken against their will.
4. The offending behaviour was sufficiently reckless or heedless of potentially lethal consequence, regardless of intent.
5. The offending behaviour was purposely to support a potentially lethal outcome.
6. Any sexual offence involving physical contact between perpetrator and victim.
7. Any offence where children were made to watch sexual activity.
8. Sexual exploitation of children through pornography or prostitution.
9. Cruelty to children including neglect.
10. Aggravated burglary in the home.

The RSR score incorporates offences under all of these criteria other than child cruelty, as actuarial scores are an ineffective means of identifying potential child cruelty reoffenders. Also, offenders who present risks only under criterion 5 (e.g. training for terrorism) are better identified through the Public Interest process.

How does the RSR score predict reoffending?

The RSR score is comprised of three sub-scores, one for contact sexual reoffending, one for indecent image reoffending and one for non-sexual violence. The contact sexual and indecent image components are only scored for men with a history of sexual offending, whilst all offenders receive a score on the non-sexual violence component, including those with no history of non-sexual violence.

Contact sexual reoffending risk

The contact sexual reoffending sub-score is based on the OASys Sexual reoffending Predictor (OSP) – see Chapter 10 – which can be calculated from limited, quickly-assessed information. The primary risk factors are sexual offending history – including several different types of sexual offence – and age. These risk factors are used to generate a 32-point score, which is used to provide estimated one- and two-year reoffending probabilities.

OSP is valid for male sex offenders. For female sex offenders, very limited actuarial sexual risk information is available, so all receive the same score.

Indecent image reoffending risk

The indecent image sub-score is based solely on sexual offence history, using risk factors already collected for OSP. History of indecent image offending is most predictive of proven indecent image reoffending, followed by history of contact sexual offences with child victims.

Non-sexual violence risk

This sub-score is more complicated, involving a larger number of risk factors. As with the existing OGRS score (see Chapter 8), some mathematical adjustments are made to the factors which assessors have inputted (e.g. to measure criminal career intensity with the 'Copas rate'), before a logistic regression model is applied to estimate reoffending probabilities. Age, general and violent criminal history, and gender are all major risk factors for serious non-sexual violent reoffending.

Two versions of the sub-score exist: the 'brief static' version uses only static risk factors (age, gender, current offence and criminal history), while the 'extended static and dynamic' version adds dynamic risk factors (OASys items similar to those scored in the OASys Violence Predictor (OVP) – see Chapter 9) and further questions about criminal history. While the static version predicts serious non-sexual violence well, the extended version is an even better predictor and therefore should be used when practical circumstances permit.

Both versions deliver the same mean score across a large sample of offenders. An individual offender will receive a higher score on the extended version than the static version if they have several dynamic risk factors for violence and/or a history of serious violent offending, but a lower score if they have few dynamic risk factors and little/no history of serious violence.

Appendix I:
Logistic regression results and 100-point scales for OGP2 and OVP2

This appendix reports the results of logistic regression models of proven general and violent reoffending over the next-two-year follow-up period. It also sets out the process by which these models were transformed into 100-point scores. Table I1 summarises the contents of the various candidate models: the two with-OGRS model types, and the finally selected no-OGRS models. Table I2 gives the logistic regression parameters for OGP2 and OVP2 respectively, while Tables I3 and I4 illustrate the scaling process. Table I5 provides assurance that the transformation process had little impact on predictive validity.

Determining how to model static and offence-free time risk

The process which led to no-OGRS models being preferred to with-OGRS models is described in Chapter 9. The dynamic factors included in each model are summarised in Table I1. Several other items were trialled but not found to be significant in any of the six models.

Table I1: Dynamic risk factors in three OGP2 models and three OVP2 models

	OGP2			OVP2		
Risk factor	Simple OGRS	Quadratic OGRS	No OGRS	Simple OGRS	Quadratic OGRS	No OGRS
OGRS4 [G or V]	√	√		√	√	
OGRS4, squared		√			√	
Bespoke static score			√			√
Offence-free months			√			√
3.4 Suitability of accommodation	√	√	√	√	√	√
4.2 Employment status	√	√	√	√	√	√
4.7 Reading/writing/ numeracy skills				√	√	
Partner score (see Note)	√	√	√	√	√	√
7.2 Regular activities encourage offending	√	√	√			
8.1 Main drug used	√	√	√	√	√	√
8.4 Frequency of main drug use	√	√	√			
9.1 and 9.2 Current alcohol use and binge drinking	√	√	√	√	√	√
9.4 History of alcohol misuse				√		
11.2 Impulsivity	√	√	√	√	√	√
11.4 Temper control	√		√	√	√	√
11.6 Problem solving skills	√	√	√	√	√	√
12.1 Pro-criminal attitudes	√	√	√	√	√	

Note. Partner score has a 0 – 4 range, and equals current relationship with partner (item 6.4) score, plus 1 point for being a domestic violence perpetrator (item 6.7), plus 3.2, where 3.2 is coded: does live with partner = 1; does not live with partner = 0. Statistically significant negative terms were suppressed when selected, and models rerun without them. 9.4 entered the simple with-OGRS OVP models with statistically significant but very small coefficients, which did not improve model fit and were difficult to incorporate into 100-point scales.

Creating the 100-point scales

To transform the logistic regression parameters into scores out of 100, the minimum and maximum feasible scores based on the logistic regression results were calculated. Given the complexity of some static risk factors, it was recognised that scores for these individual risk factors would not be comprehensible to assessors, so total static scores were produced instead by summing the individual risk factors.

The ranges between the scores associated with the 1st and 99th percentiles of the sum of the static factors, offence-free time and each dynamic risk factor ('truncated ranges') were then calculated (see Tables I3 and I4).[208] These truncated ranges for each risk factor were summed, to create total score ranges. The truncated ranges for each risk factor were then expressed as hundredths of these total ranges. For example, in OGP2, question 7.2 had a range of 0.29, which was four-hundredths of the total range (0.29 / 7.23 = 0.04), leading to an initial weighting of four points for this risk factor.

These scores were then amended to make them easier for practitioners to understand and to maximise continuity with OGP1 and OVP1. In OGP1 and OVP1, static risk factors carry 60 points and dynamic risk factors carry 40 points. In OGP2, it proved possible to allocate 15 points to offence-free time and 35 points to dynamic risk factors, with the remaining 50 points for static risk factors. In OVP2, these allocations were 10, 30 and 60 points respectively. These risk factor weightings were found to assist with simplicity, allowing many 0/2 or 0/1/2 items to have two or four point weightings.

The effects of these amendments on predictive validity were checked using the next-two-year validation sample, and are displayed in Table I5. Compared with the raw results in Tables I1 and I3, the revisions reduced predictive validity very slightly. Higher point allocations for offence-free time and dynamic risk factors reduced predictive validity to a greater extent and were therefore not selected.

Table I6 illustrates the predicted probabilities of any and violent reoffending associated with 100-point scores on the relevant predictors. Score changes of a small number of points lead to the greatest changes in predicted probabilities of any or violent reoffending (changes of 6 to 10% for a 5-point change) when OGP2 scores are between 35 and 75 or OVP2 scores are at least 55.

Following the transformation of the model results to 100-point scores, further logistic regression models were run to fit the 100-point scores to next-two-year reoffending outcomes.

[208] Steyerberg (2009) recommends that risk distributions are truncated where there is concern that extreme values could cause distortion. This would occur with both predictors' static risk factor total where the total range is at least 1.5 times the length of the truncated range. Using the maximum score would thus result in the static scale being given a large points weighting even though much of this points range would seldom be used.

Table I2: Logistic regression models of proven reoffending within the next two years: results for all and violent reoffending

OASys question / other covariate	Value/ transformation	Outcome: General B	SE(B)	Exp(B)	Violent B	SE(B)	Exp(B)
Intercept		6.7773	0.6711		5.5873	0.8296	
Age at index date, if female	Simple	-0.03027	0.2817		-0.3658	0.4278	
	Squared	0.000895	0.0121		0.0142	0.0191	
	Cubed	-0.0000131	0.000221		-0.000225	0.000363	
	To power 4	-2.276E-8	1.452E-6		1.089E-6	2.492E-6	
Female		-5.8665	2.4336		-3.3229	3.5363	
Age at index date, if male	Simple	-0.5119	0.0742		-0.5804	0.0952	
	Squared	0.01595	0.00298		0.01888	0.0039	
	Cubed	-0.000224	0.000050		-0.000273	0.000069	
	To power 4	1.130E-6	3.045E-7		1.395E-6	2.492E-7	
Offence category	Absconding/bail	0.2256	0.2046	1.974	0.0135	0.2235	1.432
	Acquisitive violence	0.0758	0.0522	1.543	0.1672	0.0576	1.181
	Burglary (domestic)	0.2621	0.0475	2.316	0.1170	0.0533	1.421
	Burglary (other)	0.2031	0.0583	2.491	0.1572	0.0644	1.678
	Criminal damage	0.0727	0.0499	1.805	0.2473	0.0538	1.596
	Drink driving	-0.1053	0.0363	1.507	0.0818	0.0441	1.359
	Drug imp./exp./product.	-0.0043	0.0544	1.597	-0.0830	0.0729	1.154
	Drug possession/supply	-0.0065	0.0363	1.604	-0.1289	0.0506	1.102
	Drunkenness	-0.1561	0.1690	1.557	0.0712	0.1773	1.450
	Fraud, forgery and misrepresentation	-0.1486	0.0492	1.401	-0.1451	0.0680	1.059
	Handling stolen goods	0.1766	0.0782	1.883	0.1212	0.0915	1.594
	Motoring (not drink driving)	0.0654	0.0354	1.794	0.1269	0.0428	1.371
	Other	-0.1523	0.0576	1.533	-0.0483	0.0688	1.314
	Public order, harassment	0.0215	0.0353	1.707	0.2099	0.0399	1.461
	Sexual (not against children)	-0.0784	0.0779	1.331	-0.2940	0.1079	0.847
	Sexual (against children)	-0.2807	0.0804	1.117	-0.9209	0.1465	0.470
	Theft	0.2303	0.0365	2.301	0.0471	0.0426	1.479
	Violence against the person	-0.0841	0.0257	1.508	0.0966	0.0303	1.285
	Vehicle-related theft	0.0847	0.0550	1.913	0.1449	0.0602	1.468
	Welfare fraud [RC]	-0.3953	n/a		0.0184	n/a	
Sanction count	1st	-2.6639	0.0951		-1.6257	0.0881	
	2nd	-2.1101	0.0913		-1.1230	0.0818	
	Per sanction	-0.0012	0.00305		-0.0117	0.00316	
Years since... (2nd only), if female	Per year	-0.0262	0.0167		-0.0593	0.0264	
Years between 1st and index sanction (2nd sanction only), if male	Per year	-0.0121	0.0065		-0.0200	0.0091	
Tempo (see Note) (3+ sanctions only), if female	Per unit of general Copas rate	0.7203	0.0497		0.5149	0.0658	

OASys question / other covariate	Value/ transformation	Outcome					
		General			Violent		
		B	SE(B)	Exp(B)	B	SE(B)	Exp(B)
Tempo (3+), if male	Per unit of gen. Copas	0.8017	0.0402		0.5149	0.0446	
Violent sanction count	None, if female	n/a			-0.8053	0.1273	
	None, if male	n/a			-0.4303	0.1036	
	One	n/a			-0.0345	0.0409	
	Per sanction	n/a			0.0682	0.0069	
Violent sanction tempo (see Note)	None	n/a			0.0595	0.0316	
Offence-free months	Simple	-0.0436	0.00317		-0.0330	0.00365	
	Squared	0.000370	0.000098		0.000212	0.000115	
3.4		0.0923	0.0136	1.097	0.0829	0.0151	1.086
4.2		0.0838	0.0097	1.087	0.0794	0.0113	1.083
Partner score [see Note]		0.0753	0.0098	1.078	0.0734	0.0111	1.076
7.2		0.1440	0.0166	1.155	0.0751	0.0181	1.078
8.1	Crack Amphetamine Heroin Cocaine Cannabis Methadone Any other None	0.3760 0.3464 0.2606 0.1996 0.0929 -0.0194 0.1076 Zero	0.0805 0.0922 0.0513 0.0825 0.0343 0.1120 0.0814 n/a	1.456 1.414 1.298 1.221 1.097 0.981 1.114	0.3236 0.1994 n/a 0.2327 0.0909 n/a 0.0329 Zero	0.0828 0.0968 0.0858 0.0272 0.0852 n/a	1.382 1.221 1.262 1.095 1.033
8.4	Daily	0.1378	0.0419	1.148	n/a		
	Weekly	0.0390	0.0438	1.040	n/a		
9.1 + 9.2		0.0543	0.0075	1.056	0.1336	0.0082	1.143
11.2		0.0792	0.0158	1.082	0.0747	0.0183	1.078
11.4		0.0641	0.0164	1.066	0.1426	0.0185	1.153
11.6		0.0929	0.0161	1.097	0.0697	0.0186	1.072
12.1		0.0618	0.0193	1.064	n/a		

Note. Copas rate = log ((count of all sanctions or violent sanctions)/(age at current sanction + *x* - age at first sanction)). Log is the natural logarithm (i.e. to base *e*, not base 10). In OGRS4/G, *x*=26. In OGRS4/V, *x*=12 for the general Copas rate, *x*=30 for the violent Copas rate. RC = reference category. See note to Table I1 on partner score. 'Cocaine' refers to cocaine hydrochloride, i.e. powdered cocaine. Crack cocaine is therefore not double counted. The initial run of the OVP model featured strong negative coefficients for heroin and methadone, and was therefore rerun without these drugs.

Table I3: Scaling OGP2 logistic regression results to produce a 100-point score

Risk factor	Distribution of product of regression parameters and risk factor score			Scores/100	
	1st percentile	99th percentile	Range	% of total range	Revised
Static factors	***-9.81***	***-5.95***	***3.86***	***53***	***50***
Offence-free time	***-1.09***	***0***	***1.09***	***15***	***15***
Dynamic factors					
3.2	0	0.18	0.18	3	2
4.2	0	0.17	0.17	2	2
Partner score	0	0.30	0.30	4	4
7.2	0	0.29	0.29	4	4
Main drug	0	0.38	0.38	5	7
Drug frequency	0	0.14	0.14	2	2
9.1 and 9.2	0	0.22	0.22	3	4
11.2	0	0.16	0.16	2	3
11.4	0	0.13	0.13	2	2
11.6	0	0.19	0.19	3	4
12.1	0	0.12	0.12	2	2
All dynamic factors	***0***	***2.28***	***2.28***	***32***	***35***
All risk factors	**-10.90**	**-3.67**	**7.23**	**100**	**100**

Note. For offence-free time and each dynamic factor, the 1st and 99th percentiles equalled the minimum and maximum possible scores.

Table I4: Scaling OVP2 logistic regression results to produce a 100-point score

Risk factor	Distribution of product of regression parameters and risk factor score			Scores/100	
	1st percentile	99th percentile	Range	% of total range	Revised
Static factors	***-10.25***	***-5.81***	***4.44***	***60***	***60***
Offence-free time	***-0.91***	***0***	***0.91***	***12***	***10***
Dynamic factors					
3.2	0	0.17	0.17	2	2
4.2	0	0.16	0.16	2	2
Partner score	0	0.29	0.29	4	4
Main drug	0	0.32	0.32	4	6
9.1 and 9.2	0	0.53	0.53	7	8
11.2	0	0.15	0.15	2	2
11.4	0	0.29	0.29	4	4
12.8	0	0.14	0.14	2	2
All dynamic factors	***0***	***2.05***	***2.05***	***27***	***30***
All risk factors	**-11.16**	**-3.76**	**7.40**	**99**	**100**

Note. For offence-free time and each dynamic factor, the 1st and 99th percentiles equalled the minimum and maximum possible scores.

Table I5: Area Under Curve (AUC) scores for validation sample, comparing raw regression, original and revised score models

Model	Raw logistic regression parameters	Score/100 using %s of total range	Two-year percentage using revised score/100
OGP2	0.7763	0.7764 *	0.7661 [RC]
OVP2	0.7657 ***	0.7643	0.7639 [RC]

Note. N = 36,221; RC = reference category. *: *p*<.05. **: *p*<.01. ***: *p*<.001.

Table I6: Association between 100-point score and predicted probabilities of proven next-two-year reoffending

Score	Predictor and probability of reoffending	
	OGP2: Any	OVP2: Violent
0	2%	0%
5	2%	1%
10	3%	1%
15	5%	1%
20	7%	2%
25	9%	3%
30	13%	4%
35	18%	5%
40	24%	7%
45	31%	11%
50	40%	15%
55	49%	20%
60	58%	27%
65	67%	35%
70	74%	44%
75	81%	54%
80	86%	63%
85	90%	71%
90	93%	78%
95	95%	84%

Appendix J:
OASys questions included within revised criminogenic need scales, OGP2 and/or OVP2 (OASys sections 3 to 12)

Table J1: Scored OASys questions by section (revised criminogenic need scales and OGP/OVP2)

Scored OASys question	Criminogenic need scales (revised)	OGP2	OVP2
Accommodation			
3.2 Accommodation status			
3.3: Currently of no fixed abode or in transient accommodation	✓		
3.4: Suitability of accommodation	✓	✓	✓
3.5: Permanence of accommodation	✓		
3.6: Suitability of location of accommodation	✓		
Employment, training and employability			
4.2: Is the person unemployed, or will be unemployed on release	✓	✓	✓
4.3: Employment history	✓		
4.4: Work related skills	✓		
4.5: Attitude to employment	✓		
4.6: School attendance			
4.7: Has problems with reading, writing and/or numeracy			
4.8: Has learning difficulties			
4.9: Any educational or formal professional/vocational qualifications			
4.10: Attitude to education/training			
Relationships			
6.1: Current relationship with close family members	✓		
6.3: Experience of childhood			
6.4: Current relationship with partner	✓	✓	✓
6.6: Previous experience of close relationships	✓		
6.7 Domestic violence: perpetrator and/or victim	✓	✓	✓
6.8 Current relationship status			
6.9 Parental responsibilities			
6.10 Parental responsibilities – are these a problem?			
Lifestyle and associates			
7.2: Regular activities encourage offending	✓	✓	
7.3: Easily influenced by criminal associates	✓		
7.4: Manipulative/predatory lifestyle			
7.5: Recklessness and risk-taking behaviour	✓		

Scored OASys question	Criminogenic need scales (revised)	OGP2	OVP2
Drug misuse			
8.1: Drugs ever misused (with full drugs breakdown)		✓	✓
8.4: Current drug noted	✓	✓	
8.5: Level of use of main drug	✓		
8.6: Ever injected drugs			
8.8: Motivation to tackle drug misuse	✓		
8.9: Drug use and obtaining drugs a major activity/ occupation	✓		
Alcohol misuse			
9.1: Is current use a problem	✓	✓	✓
9.2: Binge drinking or excessive use of alcohol in last 6 months	✓	✓	✓
9.3: Frequency and level of alcohol misuse in the past	✓		
9.4: Violent behaviour related to alcohol use at any time			
9.5: Motivation to tackle alcohol misuse	✓		
Thinking and behaviour			
11.1: Level of interpersonal skills			
11.2: Impulsivity	✓	✓	✓
11.3: Aggressive/controlling behaviour			
11.4: Temper control		✓	✓
11.5: Ability to recognise problems	✓		
11.6: Problem solving skills	✓	✓	✓
11.7: Awareness of consequences	✓		
11.8 Achieves goals			
11.9: Understands other people's views			
11.10: Concrete/abstract thinking			
Attitudes			
12.1: Pro-criminal attitudes	✓	✓	✓
12.3: Attitude towards staff			
12.4: Attitude towards supervision/licence	✓		
12.5: Attitude to community/society	✓		
12.6: Does the offender understand their motivation for offending			
12.8: Motivation	✓		

Appendix K:
Screening and quality assessment of studies

Table K1 below sets out the questions which were used for screening the study extracts and full texts. Those studies meeting the inclusion criteria were then assessed for their overall quality, with a quality assessment checklist being devised to review the full texts. As shown by Table K2, each study was rated as high, mid or low quality across four dimensions. Table K3 sets out the questions considered within each dimension. The numbers in the left hand column of Table K3 correspond to the numbers within Table K4 which presents the question responses for the 32 studies in the final review.

Table K1: Screening inclusion criteria

				Notes
1.	Was the study published in 2000 onwards?	YES/ UNCLEAR – go to Q2	NO – exclude **1_EX Date**	
2.	Is the study published in English?	YES/ UNCLEAR – go to Q3	NO – exclude **2_EX Language**	
3.	Does the study report on dynamic risk or protective factors associated with general or violent reoffending in adults?	YES/ UNCLEAR – go to Q4	NO – exclude **3_EX Topic**	Exclude studies on risk factors associated solely with sex offences, domestic violence, terrorist offences, offending in mentally ill or mentally incapacitated people. Exclude studies in which the offenders were assessed as juveniles/young offenders (aged under 18). Exclude studies on clinical outcomes and risk of serious harm to self and others.
4.	Does the study report empirical/ primary research, or is it a systematic review of empirical research?	YES/ UNCLEAR – go to Q5	NO – exclude **4_ EX Empirical**	Exclude opinion pieces etc. Include systematic reviews and meta-analyses (disaggregate and use as sources of other studies)
5.	Does the study report objective data and official outcomes?	YES/ UNCLEAR – 6_IN General 7_IN Violent	NO – exclude **5_EX Outcome**	Include arrests, convictions, parole violations; exclude self-reported outcomes

NB: For cases where inclusion is unclear, code as Q_**QUERY** and save to discuss with screening team.

Table K2: Quality assessment dimensions

	Reporting transparency	*Appropriateness of study design*	*Quality of execution*	*Relevance*
High	The aims of the study are clearly stated; information about methods and participants is complete; analytical strategy is made explicit.	The methods and sampling strategy used to answer the research question are adequate.	The methodological strategy is soundly carried out.	The population and topic under investigation are relevant to the aim of the review.
Mid	Some of the above are missing, but the study is still broadly transparent and could be replicated.	The methods and sampling strategy appears adequate, but there is some missing information.	The study is sufficiently reliable, although there are some quality issues.	The study addresses the topic in a way or in a context that is partially relevant to this review's research question.
Low	Most of the above are missing, severely limiting the possibility of evaluating the study. This necessarily has a negative impact for the rest of the appraisal.	There is a serious mismatch between the aims and the methods or no information is provided.	There are serious flaws in the execution, or not enough information is provided.	The focus or the approach of the study is only minimally relevant to the review.

Table K3: Quality assessment questions by dimension

Reporting transparency	
1	The study has clear aims
2	The methodology is clearly described
3	Primary and secondary outcomes are clearly defined and data reported for all outcomes
4	All participants are accounted for in the data reporting
5	The conclusions of the paper accurately reflect the results presented
Appropriateness of study design	
6	The methodology is appropriate to meet the aims of the study and identify the desired outcomes
7	The sampling strategy was designed to minimise bias
8	The sampling process included a representative sample of the population of ultimate interest
9	Follow up of participants was of an appropriate duration to identify the specified outcomes
10	Does the study include a range of risk factors for reoffending?
11	The statistical analysis uses appropriate tests for the data and outcomes
	For comparative studies:
12	A sample size calculation was done and is reported (where appropriate)
13	The sample size is sufficient to detect a real difference between comparison groups
14	Allocation to intervention groups was concealed and carried out such as to minimise bias
15	Allocation to intervention groups was randomised, using a method that minimises bias
Quality of execution	
16	The researchers adhered to their protocol methodology and any divergence is clearly explained
17	The planned sample size was met
18	Assessment of outcomes was done by a researcher who was blinded to the intervention group (where appropriate)
19	The statistical analysis includes a measure of confidence (P values, confidence intervals), where appropriate
Relevance	
20	The study is relevant to the adult UK population
21	The study is focused on risk factors for general or violent reoffending

Table K4: Quality assessment responses for each study

	Reporting transparency						Appropriateness of study design											Quality of execution					Relevance			Overall assessment
	1	2	3	4	5	Total	6	7	8	9	10	11	12	13	14	15	Total	16	17	18	19	Total	20	21	Total	
Allan and Dawson (2002)	Y	Y	Y	PA	Y	**H**	Y	PA	PA	PA	Y	Y	NA	NA	NA	NA	**M**	Y	PA	NA	Y	**H**	PA	Y	**M**	**M**
Brennan *et al*. (2009)	Y	Y	Y	Y	Y	**H**	Y	UC	UC	PA	Y	Y	NA	NA	NA	NA	**M**	Y	UC	NA	Y	**H**	PA	Y	**M**	**M**
Brown *et al*. (2009)	Y	Y	Y	PA	Y	**H**	Y	UC	UC	UC	Y	Y	NA	NA	NA	NA	**M**	Y	UC	NA	Y	**H**	PA	Y	**M**	**M**
Capaldi *et al*. (2008)	Y	Y	Y	Y	PA	**H**	Y	UC	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	PA	**M**	PA	Y	**M**	**M**
Coid *et al*. (2011)	Y	Y	Y	UC	Y	**H**	Y	Y	Y	Y	PA	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	Y	PA	**M**	**M**
Craig *et al*. (2004)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	PA	Y	**M**	**M**
Davies and Dedel (2006)	Y	Y	Y	Y	Y	**H**	Y	PA	Y	PA	Y	Y	NA	NA	NA	NA	**H**	Y	UC	NA	Y	**H**	PA	Y	**M**	**M**
Farrington *et al*. (2009)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	Y	Y	**H**	Y	Y	**H**	**H**
Hsu *et al*. (2009)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	PA	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	PA	Y	**M**	**M**
Kim *et al*. (2008)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	PA	Y	Y	NA	NA	NA	NA	**H**	UC	UC	NA	Y	**M**	PA	Y	**M**	**M**
Larney and Martire (2010)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	PA	NA	Y	**M**	PA	Y	**M**	**M**

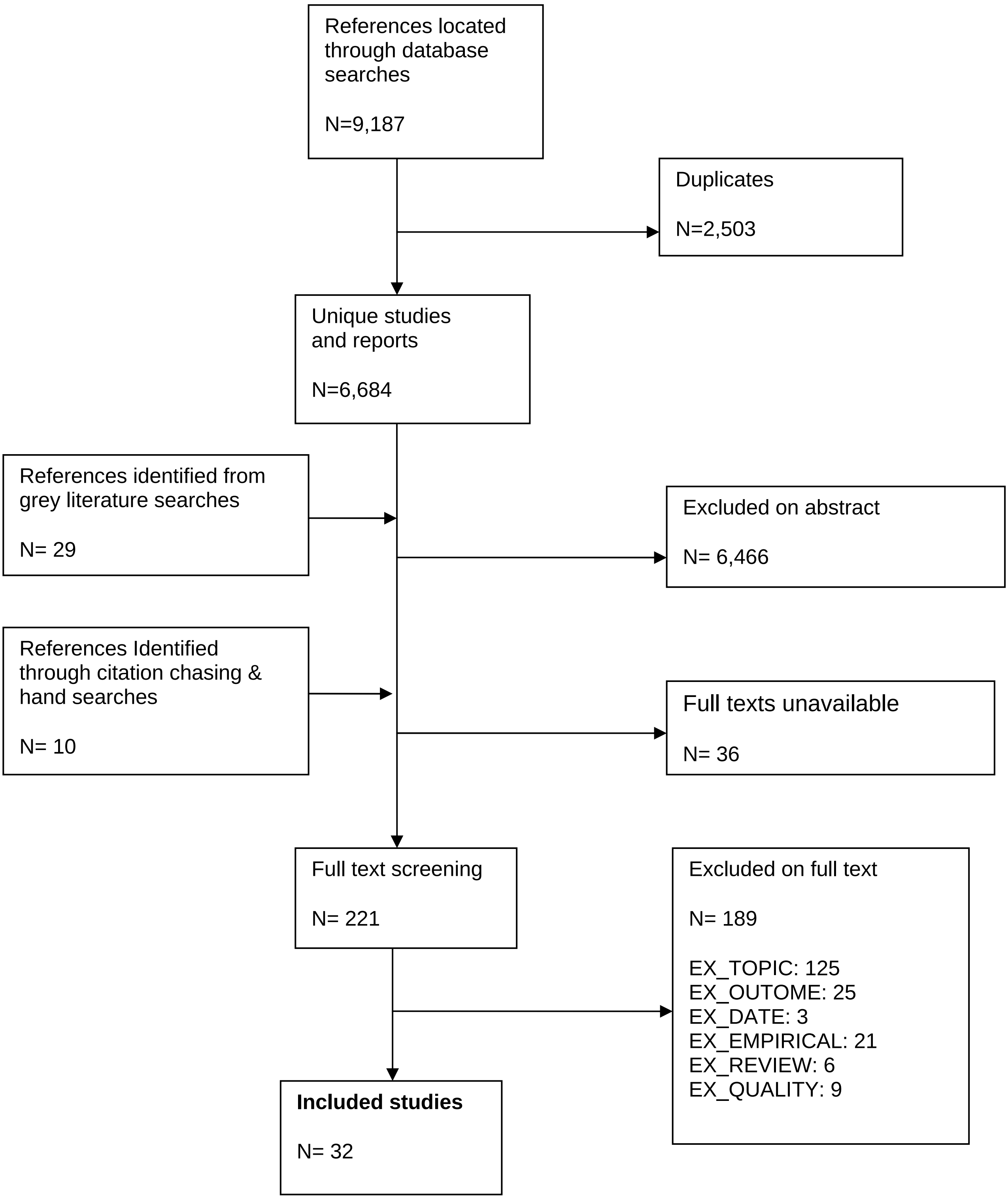
	Reporting transparency						Appropriateness of study design											Quality of execution					Relevance			Overall assessment
	1	2	3	4	5	Total	6	7	8	9	10	11	12	13	14	15	Total	16	17	18	19	Total	20	21	Total	
Lemke (2009)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	Y	Y	NA	Y	**H**	UC	Y	**M**	**M**
Listwan *et al*. (2007)	Y	Y	Y	PA	Y	**H**	Y	UC	UC	Y	Y	Y	NA	NA	NA	NA	**M**	UC	UC	NA	Y	**M**	PA	Y	**M**	**M**
Liu *et al*. (2011)	Y	Y	Y	Y	Y	**H**	Y	UC	Y	Y	Y	Y	NA	NA	NA	NA	**H**	Y	Y	NA	Y	**H**	Y	Y	**H**	**H**
Lowenkamp and Bechtel (2007)	Y	Y	Y	Y	Y	**H**	Y	UC	UC	PA	Y	Y	NA	NA	NA	NA	**M**	Y	UC	NA	Y	**M**	PA	Y	**M**	**M**
Lowenkamp *et al*. (2008)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	PA	YA	**M**	**M**
Manchak *et al*. (2009)	Y	Y	Y	Y	Y	**H**	Y	UC	UC	Y	Y	Y	NA	NA	NA	NA	**M**	UC	NA	NA	Y	**M**	PA	Y	**M**	**M**
May *et al*. (2008)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	PA	Y	**M**	**M**
Mills *et al*. (2004)	Y	Y	Y	Y	Y	**H**	Y	UC	UC	Y	Y	Y	NA	NA	NA	NA	**M**	UC	UC	NA	Y	**M**	PA	Y	**M**	**M**
Ostrom *et al*. (2002)	Y	PA	Y	Y	Y	**H**	PA	UC	Y	Y	Y	Y	NA	NA	NA	NA	**M**	UC	Y	NA	UC	**M**	PA	Y	**M**	**M**
Piquero *et al*. (2001)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	Y	Y	NA	Y	**H**	Y	Y	**H**	**H**
Raynor *et al*. (2000)	Y	Y	Y	Y	Y	**H**	Y	UC	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	Y	Y	**H**	**H**
Reisig *et al*. (2006)	Y	Y	Y	Y	Y	**H**	Y	UC	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	PA	Y	**M**	**M**

	Reporting transparency						Appropriateness of study design											Quality of execution					Relevance			Overall assessment
	1	**2**	**3**	**4**	**5**	**Total**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**Total**	**16**	**17**	**18**	**19**	**Total**	**20**	**21**	**Total**	
Salisbury *et al*. (2009)	Y	PA	Y	PA	PA	**M**	Y	UC	UC	PA	Y	Y	NA	NA	NA	NA	**M**	UC	UC	Y	Y	**H**	PA	Y	**M**	**M**
Siddiqi (2006)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	Y	Y	NA	NA	**H**	Y	Y	**H**	**H**
Stalans *et al* (2004)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	Y	Y	NA	Y	**H**	Y	Y	**H**	**H**
Tikkanen *et al*. (2010)	Y	Y	Y	Y	Y	**H**	Y	UC	UC	Y	PA	Y	NA	NA	NA	NA	**M**	Y	UC	NA	Y	**H**	PA	Y	**M**	**M**
Ullrich and Coid (2011)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	Y	Y	**H**	**H**
van Voorhis *et al*. (2010)	Y	Y	Y	Y	Y	**H**	Y	UC	PA	PA	Y	Y	NA	NA	NA	NA	**M**	Y	UC	NA	Y	**H**	PA	Y	**M**	**M**
Winters and Hayes (2001)	Y	Y	Y	Y	Y	**H**	Y	Y	Y	Y	Y	Y	NA	NA	NA	NA	**H**	UC	Y	NA	Y	**H**	PA	Y	**M**	**M**
Wormith *et al*. (2007)	Y	Y	Y	Y	Y	**H**	Y	UC	UC	Y	Y	Y	NA	NA	NA	NA	**H**	UC	UC	NA	Y	**H**	PA	Y	**M**	**M**
Yahner and Visher (2008)	Y	Y	Y	Y	Y	**H**	Y	UC	Y	Y	Y	Y	NA	NA	NA	NA	**H**	Y	Y	NA	UC	**M**	Y	Y	**H**	**M**

Y = Yes, fully; PA = Yes, partly; UA = No/unclear; NA = Not applicable.

Appendix L:
Flow of literature

Figure L1: Flow of literature



Appendix M:
Data extraction template

The template which was used for extracting data from the included studies is set out below. Data extraction for all studies was carried out by one reviewer and checked by another, with a minimum of 10% of studies independently data-extracted by two reviewers, and any differences resolved by discussion and reference to a third reviewer where necessary. It was deemed particularly important to ensure that all data were extracted on the specific measures used (predictors and outcome measures), the targeting of the assessment tools or interventions, the demographics of the sample, and any subgroup analyses completed.

Data extraction template

Study	Population	Study methods	Results	Limitations
Study id: **Study aim(s):** **Quality assessment**:	**Population**: **Country**: **Sample size**: **Reoffending category**: **Reoffending measure**:	**Study design**: **Sample selection**: **Data collection**: **Risk assessment tool**:	**Dynamic risk factors identified**: **Assessment Predictors**: **Protective factors identified**: **Outcome measures**: **Summary of main findings**:	

Appendix N:
Risk assessment instruments and other tools/scales

Instrument/tool/scale	Description
Assessment, Case Management and Evaluation (ACE)	ACE is an assessment and evaluation tool developed in partnership between the Probation Studies Unit at the University of Oxford and Warwickshire Probation Service. It assesses the criminogenic needs of offenders in a comprehensive and consistent way, assists in planning supervision to target appropriate needs, monitors progress, and evaluates to what extent supervision has addressed targeted needs. The ACE Offending Related Score is made up of 33 items grouped into the following 11 components: 1. Accommodation and neighbourhood 2. Employment, training and education 3. Finances 4. Family/personal relationships 5. Substance abuse and addictions 6. Health 7. Personal skills 8. Individual characteristics 9. Lifestyle and associates 10. Attitudes 11. Motivation
Childhood Adolescent Taxon Scale – Self Report (CAT-SR)	The CAT-SR is the self-report version of CAT which was developed by Harris, Rice, and Quinsey (1994) through their work examining the taxonomic or categorical versus dimensional nature of psychopathy. The eight items address early behaviour problems and instability in early home life (e.g. arrests prior to age 16, school suspension, alcohol abuse).
Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)	An automated decision-support software package that integrates risk and needs assessment with several other domains, including sentencing decisions, treatment and case management, and recidivism outcomes. The core assessment contains both static and dynamic factors. The latter are as follows: cognitive-behavioral, criminal associates/peers, criminal involvement, criminal opportunity, criminal personality, criminal thinking (self-report), current violence, family criminality, financial problems, history of non-compliance, history of violence, leisure/boredom, residential instability, social adjustment, social environment, social isolation, socialisation failure, substance abuse, vocation/education.
Criminogenic Needs Inventory (CNI)	The CNI was developed by the New Zealand Department of Corrections as a tool to identify the criminogenic needs of the New Zealand offending population. It is offence focused and includes the assessment of psychological needs, responsivity factors linked to offending, the role of culture in the offending period and the pre-disposing period.
DSM-III Antisocial Personality Disorder (APD) criteria	A psychiatric diagnosis of APD as defined by the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, third edition. The definition includes 12 early behavioural symptoms.

Instrument/tool/scale	Description
Historical-Clinical-Risk Management-20 (HCR-20)	The HCR-20 is a 20-item checklist to assess the risk for future violent behaviour in criminal and psychiatric populations. The items in the *history* part (H) refer to past misconduct (e.g. substance abuse or early start of violent criminal behaviour). The section *clinical* (C) scrutinises the adaptation to the current situation in treatment or detention (e.g. impulsiveness or noncompliance). *Risk management* (R) contains items that denote likely stress factors in case of release from custody.
Level of Service Inventory-Revised (LSI-R)	An actuarial scale that combines risk of reoffending and needs assessment. Developed in Canada to match offenders with probation interventions and thereby inform supervision plans. It is completed through file review and interview. A 54-item structured interview is used to systematically assess risk and coordinate information relevant to offender treatment and management planning and the determination of levels of freedom and supervision. The 54 items are grouped as follows: 1. Criminal History (10 items) 2. Education and Employment (10 items) 3. Financial (2 items) 4. Family and Marital (4 items) 5. Accommodations (3 items) 6. Leisure and Recreation (2 items) 7. Companions (5 items) 8. Alcohol/Drug Problems (9 items) 9. Emotional/Personal (5 items) 10. Attitudes/Orientation (4 items)
Level of Service/Case Management Inventory (LS/CMI)	This is a ‘fourth generation’ risk need assessment tool that seeks to structure decision making in adherence with the principles of risk need and responsivity. The LS/CMI refined and combined content of the LSI-R into 8 factors, represented by 43 items. These are grouped as follows: 1. Criminal History (8 items) 2. Education/Employment (9 items) 3. Family/Marital (4 items) 4. Leisure/Recreation (2 items) 5. Companions (4 items) 6. Alcohol/Drug Problems (8 items) 7. Procriminal Attitude/Orientation (4 items) 8. Antisocial Pattern (4 items)
MacAndrews Alcoholism Scale (MAC) scale of the Minnesota Multiphasic Personality Inventory (MMPI)	One of three substance abuse scales embedded within MMPI designed to assess the extent to which a client admits to or is prone to abusing substances.
Measures of Criminal Attitudes and Associates (MCAA)	MCAA is a two-part self report measure of criminal thinking style and associations with criminal friends. Part A intends to quantify criminal associations, Part B is a 46-item measure of criminal thinking style (criminal attitudes) including four sub-scales: violence, entitlement, antisocial intent and associates.
Michigan Alcoholism Screening Test (MAST)	The measure is a 25-item questionnaire designed to provide rapid and effective screening for life-time alcohol related problems and alcoholism.
Minnesota Multiphasic Personality Inventory (MMPI)	Personality test used in mental health. The test is used by trained professionals to assist in identifying personality structure and psychopathology.

Instrument/tool/scale	Description
Offender Assessment System (OASys)	OASys is the national risk and need assessment tool for adult offenders in England and Wales. It is used to measure an offender's likelihood of further offending; to identify any risk of serious harm issues; to develop an offending-related needs profile; to develop individualised sentence plans and risk management plans; and to measure progress and change over time.
Offender Group Reconviction Scale (OGRS)	OGRS is a static risk assessment tool, which predicts proven reoffending using actuarial scoring rules on the basis of the offender's age, gender, current offence type and criminal history. The third iteration of the tool (OGRS3) is now in operational use.
Ohio Risk Assessment System (ORAS)	ORAS was developed as a State-wide system to assess the risk and needs of Ohio offenders in order to improve consistency and facilitate communication across criminal justice agencies. The tool to be administered depends upon the decision point in the criminal justice system, but all versions contain both static and dynamic factors. For example, the Community Supervision Tool includes 35 items to assess criminal history (6 items), education, employment, & financial situation (6 items), family & social support (5 items), neighborhood problems (2 items), substance use (5 items), peer associations (4 items), and criminal attitudes & behavioral patterns (7 items).
Psychopathy Checklist-Revised (PCL-R)	Psycho-diagnostic tool most commonly used to assess psychopathy or antisocial tendencies in adults. Factor 1 provides an assessment of personality conceptualisation of personality. Factor 2 highlights the behavioural correlates of psychopathy as manifest in a chronically unstable antisocial lifestyle.
Risk Needs Inventory (RNI)	An adapted version of the LSI-R, used by the Queensland Department of Corrective Services (Australia). It has sub-scales covering criminal history, social interaction, driving, health, addiction problems and attitudes towards supervision.
Rutter A2 scale	Rutter parents' scale is a 31-item questionnaire that measures psychiatric symptoms and deviant behaviour. Items are grouped into three sections: eight questions address health problems, such as headache and bed-wetting; five questions address sleeping, eating, speech, and stealing; and 18 questions address other aspects of the child's behaviour.
Rutter B2 scale	The Rutter teachers' scale is a 26-item questionnaire designed to evaluate a child's behaviour at school.
Salient Factor Score	A seven item instrument to assess a prisoner's likelihood of recidivism after release. The items, which are predominantly static in nature, are: prior criminal convictions, prior criminal commitments, offender's age at current offence, length of time without incarceration, any history of escape/supervision violation, and record of opiate dependence.

Instrument/tool/scale	Description
Special Hospitals Assessment of Personality and Socialization (SHAPS) Impulsivity scale	The SHAPS is a self-report assessment consisting of 213 items divided into the following 10 scales: 1. Lie 2. Anxiety 3. Extroversion 4. Hostility 5. Introversion 6. Depression 7. Tension 8. Psychopathic deviate 9. Impulsivity 10. Aggression
Statistical Information on Recidivism Scale-R1 (SIR-R1)	A measure of demographic and criminal history characteristics that yields probability estimates of reoffending within three years of release.
Violence Risk Appraisal Guide (VRAG)	The VRAG is an actuarial risk assessment instrument, developed in Canada as an aid to predict violent recidivism. It has 12 items requiring rating of the index offence, psychopathy, alcohol use and past non-violent crime.
Virginia Risk Assessment Instrument	An objective research-based instrument that assists Pretrial Services Officers in the performance of their duties by identifying a defendant's level of risk of failure (failure to appear and/or new arrest) if released pending trial.
Wisconsin Risk/Need Instrument	A risk and need tool developed for classification purposes, assigning offenders to differing levels of service. The risk scale consists of 13 items including prior criminal convictions, prior incarceration, prior supervisions, recent employment, alcohol use, drug use, pro-criminal peers, offence severity and pro-criminal attitudes. The need scale is made up of 14 items including mental stability, family relationships, drug/alcohol use, employment, education, financial, attitude, residential stability, intelligence, health and sexual dysfunction,