Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys Violence Predictor

Howard, Philip; Dixon, Louise

DOI:
10.1037/lhb0000012

Full reference for this post print article: Howard, P. D, & Dixon, L. (2012). Identifying Change in the Likelihood of Violent Recidivism: Causal Dynamic Risk Factors in the OASys Violence Predictor. *Law and Human Behavior, 37, 163-174*

Journal homepage: http://www.springer.com/psychology/psychology+%26+law/journal/10979

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

By Philip D Howard[1] and Louise Dixon[2]

[1]National Offender Management Service, London, United Kingdom, and University of Birmingham

[2]University of Birmingham

Author Correspondence should be addressed to: Philip D. Howard, 26 Montana Close, South Croydon, Surrey, CR2 0AT United Kingdom. E-mail: philipdavidhoward@gmail.com

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

Abstract

Recent studies of multi-wave risk assessment have investigated the association between changes in risk factors and violent recidivism. This study analyzed a large multi-wave dataset of English and Welsh offenders (N = 196,493), assessed in realistic correctional conditions using the static/dynamic Offender Assessment System (OASys). It aimed to compare the predictive validity of the OASys Violence Predictor (OVP) under mandated repeated assessment and one-time initial assessment conditions. Scores on five of OVP's seven purportedly dynamic risk factors changed in 6 – 15% of pairs of successive assessments, while the other two seldom changed. Violent reoffenders had higher initial total and dynamic OVP scores than nonreoffenders, yet nonreoffenders' dynamic scores fell by significantly more between initial and final assessment. OVP scores from the current assessment achieved greater predictive validity than those from the initial assessment. Cox regression models showed that, for total OVP scores and most risk factors, both the initial score and the change in score from initial to current assessment significantly predicted reoffending. These results showed consistently that OVP includes several causal dynamic risk factors for violent recidivism, which can be measured reliably in operational settings. This adds to the evidence base which links changes in risk factors to changes in future reoffending risk, and links the use of repeated assessments to incremental improvements in predictive validity. Further research could quantify the costs and benefits of reassessment in correctional practice, study associations between treatment and dynamic risk factors, and separate the effects of improvements and deteriorations in dynamic risk.

Keywords: Criminal recidivism, dynamic risk assessment, multi-wave prediction, OASys

Introduction

The success of a correctional organization's attempts to treat and manage offenders is partly dependent on its ability to identify those offenders most at risk of recidivism. Several prominent, well-validated tools have been developed for this task. Some tools solely utilise a narrow range of static risk factors based on demographics and criminal history, and invariably use actuarial scoring (e.g., the Offender Group Reconviction Scale, Version 3 (OGRS3) (Howard, Francis, Soothill, & Humphreys, 2009), Risk Matrix 2000 (Thornton, 2007), Static-99 (Hanson & Thornton, 1999)). Other tools additionally utilise a broader range of risk factors spanning socioeconomic, interpersonal, substance misuse, mental health and/or cognitive/attitudinal domains. This latter group can either use risk factors to structure professional judgement (e.g., Historical, Clinical, Risk Management-20 (HCR-20) (Webster, Douglas, Eaves, & Hart, 1997)), or classify risk through actuarial methods (e.g., Level of Service Inventory-Revised (LSI-R) (Andrews & Bonta, 1995), Level of Service/Case Management Inventory (LS/CMI) (Andrews, Bonta, & Wormith, 2004), Violence Risk Assessment Guide (VRAG) (Quinsey, Harris, Rice, & Cormier, 1998)).

Correctional service resources are scarce, with caseloads averaging dozens of offenders per correctional officer in the US (Petersilia, 2003) and many European jurisdictions (van Kalmthout & Durnescu, 2008). The speed and reliability with which the most narrowly-focused, purely static actuarial tools can be coded is therefore valuable, and justification is needed for the use of more complex assessment methods, whether using actuarial or structured professional approaches. Including a broad spectrum of risk factors could improve predictive validity (Harris et al., 2003). It could also facilitate insight into appropriate treatment and management of each offender (Wong & Gordon, 2006), although this will only be true if offenders score highly on these risk factor scales because they have potentially dynamic needs rather than because they are members of a persistently antisocial taxon (Hanson, 2005). While strong predictive validity and treatment relevance form only part of a wider set of criteria to be fulfilled by a risk assessment procedure (Hart, 2001; Hanson & Howard, 2010), any assessment tool which requires the expense of measuring a broad range of risk factors should ideally achieve both of these goals.

This study aims to investigate whether causal dynamic risk factors are present in a very large sample of UK offender assessments completed in realistic correctional conditions, and therefore whether the assessed factors have treatment relevance. It also uses this sample to test whether using repeated rather than one-time assessments improves the predictive validity of an actuarial violence risk score created by combining static and dynamic variables.

Existing Evidence on the Utility of Measuring Dynamic Risk Factors

Existing evidence offers some support for efforts to achieve predictive and treatment targeting goals by assessing purportedly dynamic risk factors. The evidence that such assessment improves predictive validity is weaker than the evidence that it improves selection of appropriate offenders for treatment. Many comparisons of the predictive validity of multiple instruments on individual samples have been undertaken and, for more widely studied instruments, have been consolidated through meta-analyses (Campbell, French, & Gendreau, 2009; Hanson & Morton-Bourgon, 2009; Yang, Wong, & Coid, 2010). Hanson and Morton-Bourgon (2009) found modest empirical support for combining static and dynamic risk factors, while the other two meta-analyses offered less evidence that using dynamic risk factors improves predictive validity. By contrast, assessment of dynamic risk factors is a cornerstone of the risk-needs-responsivity model of offender treatment (Andrews & Bonta, 2010), in which treatment is made available to those more likely to reoffend, has the intermediate goal of addressing criminogenic needs – “dynamic risk factors that, when changed, are associated with changes in the probability of recidivism” (Andrews & Bonta, 2006, p. 49) – and delivered in a manner which suits the offender’s learning style and ability level. The proven benefits of offender treatment (McGuire, 2004) are best realized when two or ideally all three of the risk, need and responsivity principles are followed (Andrews & Bonta, 2006). As well as ensuring that treatment targets are relevant to the individual, appropriate allocation to treatment programmes has been associated with increased likelihood of programme completion (Taxman & Thanner, 2006; Palmer et al., 2009).

Key Considerations in the Definition and Measurement of Dynamic Risk Factors

The inclusion of measures of personal and social needs in risk assessment tools is therefore likely to increase the utility of those tools in planning treatment and case management, but has not yet been proven through meta-analysis to improve prediction of reoffending. Risk predictors tend to strongly weight criminal history, and the extent and nature of an offender's criminal history will have been influenced by their past personal and social risk factors (Beech & Ward, 2004). Incremental predictive validity is therefore most likely to be attained by measuring the offender's present situation on risk factors which do change over time.

Skeem and Mulvey (2002) differentiate risk status from risk state. Risk status is a label attached to the individual's static risk classification, whereas the individual's risk state can change as their dynamic risk factors change. As Douglas and Skeem (2005) describe, risk state has been conceptualized in several overlapping ways: in its relevance to interventions for violence reduction, in identifying criminogenic needs as part of the risk-needs-responsivity model, or through identifying causal dynamic risk factors. To paraphrase Kraemer et al. (1997), an item or scale is necessarily a risk factor for violent recidivism if it (1) is correlated with violent recidivism, and (2) precedes that recidivism in time. Such a risk factor will fulfil the further properties of a causal dynamic risk factor, when (3) the risk factor is capable of changing, and (4) when change in the risk factor occurs, change in the probability of violent recidivism also occurs. Dynamic risk factors, in the current paper, are those which fulfil criteria (1), (2) and (3). Dynamic risk factors may also be stable or acute – slowly or rapidly changing, respectively (Hanson &

Harris, 2000).

If causal dynamic risk factors do exist, current measures would often yield different results from measures taken when the same individual had opportunities to offend in the past. The effects of these factors on offending behaviour could not therefore be estimated fully by proxy using criminal history data. Studies are thus most likely to identify incremental predictive validity for dynamic risk factors if they incorporate repeated assessment over time, yet "the vast majority of [existing] studies have relied exclusively on 'single-wave' research designs…. [which] assess dynamic risk factors once only.… [and therefore] have treated dynamic factors as if they were static-immutable" (Brown, St. Amand, & Zamble, 2009, p. 26). Applying the risk-needs-responsivity model to guide work with offenders should similarly be most effective when timely measures of criminogenic needs are applied.

Drawing these conceptualizations together, the most useful dynamic factors to be targeted in violence risk assessment processes would be (1) acute enough that changes would often be detected over the course of an individual's contact with forensic or correctional services, (2) amenable to intervention, and (3) predictive of violence at initial assessment and when reassessments occur. This paper will evaluate whether the risk factors labelled as dynamic in an existing violence risk scale, the OASys Violence Predictor (OVP) (Howard & Dixon, 2012), meet the first and third of these requirements by fulfilling the properties of causal dynamic risk factors.

Existing Multi-wave Studies of Dynamic Risk Factors

Successfully assessing causal dynamic risk factors could therefore yield considerable benefit. Many prominent risk assessment instruments include purportedly causal dynamic risk factors (i.e., risk factors implicitly or explicitly promoted by the designers as having the causal dynamic properties outlined above), yet only a few empirical investigations of changes in these risk factors have been completed.

These studies have generated promising evidence, contrasting with the disappointing meta-analytic evidence on the predictive benefit of single-wave measurement of purportedly dynamic risk factors. Four recent studies of nonsexual violence risk and a pair on sexual risk among adult offenders are described below. Earlier studies, often exploratory and/or small-scale, were discussed by Douglas and Skeem (2005) and Hanson (2005), while Van der Put et al. (2011) studied change in adolescents' risk factors.

Brown et al. (2009) reported a highly structured study of male Canadian federal prisoners (N = 136). Five static measures and 18 dynamic measures were assessed pre-release, and the dynamic measures were also assessed one and three months after release. Repeated measures analysis and pairwise comparisons determined that the majority of dynamic risk factors changed significantly, though not always in the expected direction. Comparing the predictive accuracy of Cox survival models with time-dependent covariates, a static-only model was outperformed by a static and prospective dynamic model, which in turn was outperformed by a static and time-dependent dynamic model. These results demonstrated that combining static and dynamic risk factors was important in pre-release assessment, and that tracking changes in dynamic risk factors post-release improved prediction further.

Jones, Brown, and Zamble (2010) extended the mean followup of Brown et al. (2009) from 10 months to 6.5 years, added a 6-month post-release phase, and contrasted ratings made by graduate researchers (the Brown et al. method) with the more naturalistic ratings made by serving parole officers. Most results were similar to those of Brown et al., and the superiority of researcher ratings was not statistically significant. Jones et al. suggest that parole officers' greater contact with offenders and access to collateral information may have improved the quality of their assessments, thus compensating for the weaknesses created by training deficits and operational constraints.

Schlager and Pacheco (2011) examined change in LSI-R total and subcomponent scores among offenders under community supervision (N = 179), though their data did not include recidivism outcome.

LSI-R comprises a criminal history subcomponent (10 items) and nine subcomponents covering social/personal problems (44 items, of which some are static by definition and others are potentially dynamic). All items are scored on a binary basis, and the total score is a simple sum of the subcomponent scores. Schlager and Pacheco found that mean total scores, and mean scores for eight of the ten subcomponents, fell significantly over an interval of approximately six months.

Quinsey, Jones, Book, and Barr (2006) examined sets of monthly staff ratings for 595 forensic psychiatric patients, using the Proximal Risk Factor Scale and Problem Identification Checklist. Most subscales of these two dynamic risk scales discriminated between recidivists and nonrecidivists, as did VRAG score. For both new violent acts and other new antisocial behaviours, significant increases occurred in some but not all subscales of both dynamic risk scales in the months leading up to recidivism.

Olver, Wong, Nicolaichuk, and Gordon (2007) reported on the properties of the Violence Risk Scale – Sexual Offender version (VRS-SO). This tool is predominately dynamic, with change assessed through evidence of readiness to undertake treatment and development and maintenance of positive attitudes and coping strategies during and after treatment. Cox regression survival analyses ($N = 351$, mean 10 year follow-up) were used to examine the relationship between change on dynamic items and recidivism. Positive change was detected on all three factors (sexual deviance, criminality and treatment responsivity) derived from the 17 dynamic items, and change in total dynamic score was predictive of sexual recidivism after controlling for total pre-treatment score. Olver and Wong (2011) reported on a similar sample ($N = 321$), in which the predictive validity of Static-99 was greatest among offenders with less treatment change. Change in VRS-SO was predictive within the subgroup with high Static-99 scores.

Taken together, these five multi-wave studies tended to support the causal dynamic nature of at least some of the risk factors measured, and therefore support the utility of assessing purportedly dynamic risk factors. Risk factor scores often changed and, where a recidivism study was conducted, models including score changes outperformed those which mimicked single-wave studies by excluding these data. As would be expected given that the offenders/patients studied received treatment or other rehabilitative efforts, score changes usually indicated reductions in criminogenic need. While they covered several different risk assessment instruments and offender/patient groups, each of the studies was relatively small-scale, and some studies used ratings made by researchers rather than operational staff. Further data would help to consolidate or challenge these emerging findings.

Dynamic Risk Measurement and the Offender Assessment System

The National Offender Management Service (NOMS) – the adult correctional service of England and Wales – imposes a regime of structured risk assessment upon its constituent prisons and probation trusts. Within NOMS, OGRS3 and the Offender Assessment System (OASys; Home Office, 2006) are ubiquitous as a static predictor of general recidivism and a static/dynamic risk assessment and sentence planning tool respectively. OASys includes two actuarial risk predictors, both of which combine static and purportedly dynamic risk factors. OVP (Howard & Dixon, 2012) is the focus of this paper, and is described in Measures below. It is complemented by a risk scale for nonviolent recidivism, the OASys General reoffending Predictor (OGP) (Howard, 2009). OGP and OVP scores help assessors to determine each offender's "tier" (level of supervision) through estimation of likelihood of general recidivism and risk of serious harm (National Probation Service, 2008, includes an earlier version of these tiering rules). This tiering process is intended to be dynamic – that is, allowing offenders to move between tiers if their risk alters over the course of community supervision – and therefore relies upon a previously untested assumption that the OASys predictors include causal dynamic risk factors. The procedures used to allocate scarce places on accredited offending behaviour programmes (most recently National Offender Management Service, 2010, though updates will occur) are in the process of switching from OGRS to the OASys predictors, and again can only benefit from evidence that the dynamic elements of these predictors are causally associated with recidivism. More broadly, despite a long tradition of research indicating the importance of adhering to risk, need and responsivity principles (Hollin, 1995), NOMS has been one of several worldwide correctional services which has struggled to maintain professional and political commitment to evidence-based treatment principles (Gendreau, Smith, & Thériault, 2009). The consistent application of these principles should be aided by the ready availability of risk predictors which are proven to accurately reflect the offender's current risk status and therefore reduce the temptation to dilute the integrity of treatment allocation through unstructured clinical overrides.

This paper uses survival analytic methods to examine whether changes in OASys Violence Predictor scores occur and are associated with changes in the hazard (likelihood at a point in time) of violent recidivism. The frequency of change in each of the OASys dynamic risk factors scored in OVP is measured. The extent to which rescoring OVP through repeat OASys assessment improves OVP's predictive validity is estimated. Cox regression modelling is used to identify associations between changes in OVP's dynamic risk factors and changes in the hazard of violent recidivism.

The focus of this paper is firmly upon OVP and its constituent items and subscales as they are completed by probation officers, in the real world setting of English and Welsh correctional practice. As such, it capitalises upon the very large scale of OASys use in NOMS, and the repeat administration of OASys assessments over the course of community supervision. OASys practice is imperfect: Howard and Moore (2009) showed that assessors often failed to fulfil NOMS' mandatory National Standards on OASys frequency and quality (Ministry of Justice, 2007), with some offenders having only one recorded assessment over a lengthy supervision period and other offenders' assessment sequences showing a total absence of assessed change across a large number of dynamic risk factor items. The results of our data analyses therefore do not reveal the extent of improvements in predictive validity associated with repeated assessment in ideal conditions. Instead, as with the parole officer ratings of Jones et al. (2010), they indicate whether mandated repeated assessment improves upon the predictive validity of one-time assessment in realistic correctional conditions.

Method

Ethics

This study was approved by the Ethics Board of the University of Birmingham Department of Psychology, and through the Ministry of Justice's Research Quality Approval process. Access to Police National Computer (PNC) data, and its merging with OASys data, was granted by the Police Information Approval Panel. All approving bodies were aware that the research involved the use of nonanonymized data without seeking consent from the offender participants, as permitted by the crime reduction provisions of the Data Protection Act 1988. All data on individual offenders were always stored on secure government networks, and only accessed by NOMS and Ministry of Justice staff who have enhanced government security clearance and frequently access similar data on government business. The authors follow the codes of conduct of Government Social Research and the British Psychological Society.

Participants

All OASys assessments completed between October 2004 and March 2008 (*N* = 2,682,600) were obtained from the OASys research database (see Procedure). Assessments completed prior to this date had been eligible for the sample originally used to construct OVP and are excluded.

Initial assessments of individuals subject to pre-sentence court reports, commencing community sentences or supervision upon release from custody were systematically filtered to remove all assessments without complete data on dynamic risk factors and key variables necessary for matching, and to remove duplicate assessments relating to the same individual and sentence. (The data completeness filtering resulted in the dataset submitted to the PNC being 14% smaller than it otherwise would have been. Duplicates occasionally occurred due to difficulties with the OASys IT system. Multiple initial assessments for a single individual could be included when it could be securely determined that they related to separate sentences; we use the term ***offender*** to refer to an individual tracked across one sequence of assessments starting with an initial assessment, i.e., one individual can be multiple offenders.) Offenders with community sentences or post-custodial supervision of less than four months duration were also excluded, as they were unlikely to receive an OASys review assessment (Ministry of Justice, 2007). The remaining 199,892 initial assessments were matched with the PNC research database and NOMS recall database (see Procedure below) in July 2010. The successfully matched dataset of 196,655 initial assessments (98.4% match rate) were traced back from the conviction date to ascertain criminal history and traced forward from the sentence/release date to ascertain proven reoffending rates and check for recall to custody. Of these, 162 were excluded from further analysis because, on “day zero” of the followup, they committed violent reoffences or were recalled to custody. Neither of these groups of offenders are of interest when studying the impact of changes in risk assessment score as community supervision progresses. Therefore, the remaining 196,493 initial assessments were eligible for matching procedures to track later OASys assessments (see Procedure) and inclusion in at least some survival analyses.

Of the 196,493 cases, the mean length of followup was 27.1 months (standard deviation 14.9 months, range 15 to 57 months), 87% were male, 11% were of nonwhite ethnic origin, and the mean age was 30.7 years (standard deviation 10.4 years). They included 25% on licence from a custodial sentence, while 34% had an index offence included in OVP’s classification of violent offences (Howard & Dixon, 2011). A majority, 59%, were not in full time employment, 39% had no educational or formal professional / vocational qualifications, and 17% were of no fixed abode or living in transient accommodation.

Measures

Previous sanctions and proven reoffending.

Previous sanctions for an offence group are the number of formal criminal sanctions (convictions, cautions, reprimands and final warnings) the offender has received for that offence group up to and including the index offence. Proven reoffending comprises offences committed after the date of community sentence or release from custody and by 2 July 2009. Prior to matching with OASys data, the PNC research database had been last updated on 2 July 2010, so a ‘buffer period’ of 12 months allowed conviction to occur and data to be entered onto the PNC. The administrators of the PNC research database confirmed that very few changes to the data occur when buffer periods are extended beyond 12 months. Reoffending dates for violent offences and a subset of homicide and wounding offences were coded as outcome measures. The operationalization of these offence groups was detailed by Howard and Dixon (2011). Imprisonment dates for any reoffence were coded as censoring events.

Offender Assessment System (OASys).

The Offender Assessment System (OASys) (Home Office, 2006) is a structured clinical risk/needs assessment and management tool. It is used throughout NOMS, to inform court reports on offenders convicted awaiting sentence, and manage those serving custodial sentences of at least 12 months (which are usually partly served in the community) or noncustodial sentences involving supervision. In 2010, 65% of noncustodial sentences managed by NOMS lasted one year, and 25% lasted two years (Ministry of Justice, 2011a). Assessments are reviewed periodically over the course of the sentence. In 2010/11, approximately 860,000 assessments were completed on 360,000 offenders by 18,500 staff. OASys has strongly influenced the design of the offender assessment systems of several other European countries (van Kalmthout & Durnescu, 2008). All OASys assessors are trained in interviewing skills, how to complete OASys and use its IT application, and must follow ongoing quality assurance and countersignature procedures. Many assessors are professionally qualified Probation

Officers, while offenders anticipated to be lower risk are assessed by staff with more limited professional training.

OASys consists of four main components: an analysis of offending-related factors, a risk of serious harm analysis, a summary sheet and a sentence plan. The offending-related factors component includes 13 sections, covering criminal history, Analysis of [current] Offences, ten social/personal risk factors which may have dynamic properties (accommodation; education, training and employability; financial management and income; relationships; lifestyle and associates; drug misuse; alcohol misuse; emotional wellbeing; thinking and behaviour, and attitudes) and suitability to undertake sentence-related activities (e.g., unpaid work, offending behavior programs). During this study's 2004-08 sampling frame, each social/personal risk factor was assessed using between four and ten items, each scored on a 0/2 or 0/1/2 basis, totalling 62 items. The risk of serious harm analysis component provides a structure for clinical case formulation and Risk Management Plan for offenders considered likely to commit harmful acts in the future. The summary sheet component uses IT functionality to automatically score OVP and the OASys General reoffending Predictor, a predictor of nonviolent reoffending (Howard, 2009). The Sentence Plan combines responsivity considerations with the dynamic risk factors and risk of serious harm analysis to determine case management strategies and interventions.

Moore (2009) examined the internal reliability and construct validity of the ten social/personal risk factor sections and the criminal history section. Eight of these sections were described by single factors, but three split into two factors each and a further 'violence' factor emerged. Morton (2009) produced promising but methodologically weak inter-rater reliability results. Howard and Moore (2009) produced preliminary evidence supporting the causal dynamic nature of OASys's social/personal risk factors by comparing item and section (risk factor total) scores over series of assessments during community supervision periods of up to two years. Most item scores changed in between 5% and 20% of original/final assessment pairs, only 30% of such assessment pairs included no changes in any social/personal item score, and changes in section scores between first and second assessments were predictive of recidivism at third assessment. On this basis, the social/personal risk factors are described as dynamic for the remainder of this paper.

OVP.

The OASys Violence Predictor is an actuarial predictor of proven violent reoffending, based on static and dynamic risk factors measured within OASys. Proven violent reoffending is classified in OVP as any proven reoffending involving offence(s) of homicide and assault, threats and harassment, violent acquisitive offences (robbery and aggravated burglary), public order (e.g., affray, being drunk and disorderly in public), criminal damage and/or weapon possession. Howard and Dixon (2011) determined, from OASys and PNC data, that this classification would aid prediction of future homicide/assault and the most serious violent offence subcategory of homicide and wounding, and that maximization of predictive validity for both sexual and nonsexual violent recidivism would be aided by assessing sexual recidivism risk separately.

Howard and Dixon (2012) generated OVP's scoring system on the basis of an ordinal logistic regression model, using a construction sample of 15,918 initial assessments completed between January 2002 and September 2004. OVP scores range from 0 to 100. Sixty points are available for static risk factors: age (20 points), gender (5), and previous sanctions for violent (25) and nonviolent (10) offences. Forty points are available for selected dynamic risk factors and items: failing to recognise the impact of offending (4 points), accommodation (4), employability (6), alcohol misuse (10), current psychiatric treatment (4), temper control (6), and antisocial attitudes (6). The resultant score is translated into probabilities of proven violent reoffending within 1 and 2 years (the latter is conventionally used in NOMS) through logistic functions. Using a validation sample of 49,346 initial assessments completed between October 2004 and September 2005, Howard and Dixon (in press) found that OVP had an Area

Under Curve (AUC) of .74 for all proven violent reoffending and .72 for proven homicide and wounding reoffending. It was a significantly better predictor of both outcomes than OGRS3 and the static actuarial Risk Matrix 2000/V (Thornton, 2007).

OGRS3.

OGRS is used slightly more frequently than OASys, as it is also used for oral court reports and nonrehabilitative sentences such as Community Orders involving unpaid work. It is a purely actuarial estimate of the percentage probability of proven reoffending for most recordable offences within a two-year follow-up, combining seven criminal history and demographic variables in a logistic function. It has been periodically revised and recalibrated, and version 3 (OGRS3; Howard et al., 2009) has recently been introduced. OGRS achieved a good weighted, adjusted AUC of .71 from two violence prediction studies in Yang et al. (2010)'s meta-analysis.

Procedure

The Police National Computer (PNC) research database.

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

The OASys research database.

Data from completed assessments are copied to the OASys Data Evaluation and Analysis Team, a research and statistics office within NOMS headquarters. Data completeness and integrity checks are undertaken before producing subsets for analysis.

The NOMS recall database.

NOMS headquarters maintains information on offenders considered a potential risk to the public. This includes a dataset of recalls to custody for breach of licence conditions, including offenders' names and personal identifiers and the data of each recall.

Matching OASys, PNC and recall data, and scoring OGRS3 and OVP.

Initial offender assessment data, extracted from the OASys research database as described in Participants above, were matched with three further datasets to create the final dataset for analysis. The PNC database was used to determine dates of earliest violent reoffending and imprisonment for any reoffence, and provided static data to score OGRS3 and OVP, as these tools were not implemented in OASys until August 2009. The recall database was used to identify followup censoring in the form of earliest recall to custody date. The OASys research database provided review assessments, which contain (potentially) revised dynamic risk factor data with which to rescore OVP. Changes in OVP risk factors and total scores between initial and review assessments formed the basis of most data analyses.

The PNC records of OASys-assessed offenders were retrieved by PNC research database administrators, on the basis of offenders' name, date of birth, sex and index offence conviction date. Name, date of birth and sex were used by recall database administrators to retrieve recall data. We then merged the reoffending and recall records with the initial assessment data. Proven violent reoffending, using a cut off date of 2 July 2009, was found for 65,172 offenders, while 50,481 were imprisoned for any offence. 16,507 offenders were recalled to custody during this time. We used the OASys database's internal system identifiers to track offenders who had an initial assessment and therefore select their subsequent ***review assessments***, which contain (potentially) revised dynamic risk factor data with which to rescore OVP. Some 663,245 review assessments were completed by 1 July 2009 and therefore preceded the cutoff date for PNC proven reoffending dates. Of these, 378,596 predated the earliest of first violent reoffending, first reimprisonment for any offence and first recall to custody, while 439,204 predated the earliest of first homicide/wounding reoffending, first reimprisonment and first recall. The combined datasets of initial and review assessments therefore included 575,089 and 635,697 assessments for violent and homicide/wounding reoffending respectively. The combined dataset for violent reoffending included at least one review assessment for 146,755 (75%) of the 196,493 initial assessments, while the combined homicide/wounding dataset included review(s) for 158,659 (81%). For each reoffending outcome, the most recent assessment prior to reoffending or censoring was the offender's ***final assessment***, regardless of whether this was a review or initial assessment.

Analysis

Overview.

This overview outlines the stages of analysis reported below, and clarifies key statistical concepts. Patterns of review assessment over time are presented, checking the proportions with any reassessment, any change in each of the items constituting each of the seven dynamic risk factors in OVP separately, and any change in the items constituting the total OVP score. If offsetting increases and decreases within a risk factor, or in the total score, result in zero overall difference, this is counted as a change.. Absolute and net changes in OVP's risk factors between successive assessments are measured. Reoffenders' and nonreoffenders' initial risk factor scores and changes in scores between initial and final assessment are compared. ***Concordance Indices*** compare the predictive validity of OGRS3 and OVP scores at initial and final assessments for each outcome. To utilise data from all assessments and control for variations in time at risk following each assessment, Cox regression models are fitted. Some models utilise only the initial assessment and other models incorporate review assessments by adding initial-to-most-recent-assessment changes in dynamic risk factor scores as ***time-dependent covariates***. Models using the total OVP dynamic score and models separating its constituent risk factors are both fitted. A summary measure of the acuteness of each dynamic risk factor, combining its change frequency and regression coefficient for initial-to-most-recent changes, is presented.

Time-dependent covariates.

Time-dependent covariates are covariates whose values change over the course of the followup, i.e. because the offender has had a new OASys assessment. (We follow standard NOMS practice by not recalculating static risk factor scores mid-followup due to ageing). Time-dependent covariates were incorporated into Cox analyses by splitting the followup into the periods between OASys assessments. For example, an offender who was reassessed 90 days after their initial assessment and reoffended after 120 days was included twice in the Cox sample: for the 0 to 90 day period, with their initial scores, and for the 90 to 120 day period, with scores from their reassessment, described as the ***current scores*** in the Results (which, for any given individual, might be the same as or different from their initial scores).

Concordance Indices.

The Concordance Index (C; Harrell, Lee, & Mark, 1996; Kattan, 2003) is a measure of predictive validity which can be used with time-dependent covariates and when periods at risk vary. C 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 combining two sets of comparisons: (a) every pair of reoffenders and nonreoffenders, ensuring that the nonreoffender was at-risk for at least as long as the reoffender (if the nonreoffender was at risk for a shorter period, we cannot be sure that they would not have reoffended if given more time to do so), and (b) every pair of reoffenders except pairs who reoffended on the same followup day.

Comparison (a) is the basis of the more familiar AUC measure, but comparison (b) differentiates C from AUC by also checking whether earlier reoffenders had higher risk predictor scores than later reoffenders. As this is more difficult than merely predicting whether or not reoffending will occur at all, C scores are lower than AUC scores for the same sample. C scores for different outcomes can only be compared with caution, as the ratio of 'easy' yes/no comparisons to 'hard' earlier/later comparisons is greater for less frequent (e.g., homicide/wounding) outcomes. With time-dependent covariates, the standard calculation method described above is varied by using each offender's predictor scores in effect on the reoffender's day of reoffending ('current' scores), rather than using each offender's initial scores. (For example, if reoffender X was not reassessed before reoffending 90 days after his initial assessment, and nonreoffender Y was reassessed 70 and 130 days after his initial assessment, scores from X's initial assessment and Y's 70-day reassessment are compared when calculating C.)

Confidence intervals for C cannot be calculated for large samples, as resampling methods such as the bootstrap (Harrell et al., 1996) impose impractical computational demands. Highly complex alternative processes trialled in machine learning research (Rayker, Steck, Krishnapuram, Dehing-Oberije, & Lambin, 2007) produce narrow confidence intervals on samples in the low hundreds, suggesting strongly that the magnitude of our predictors' C differences (see Results) represent real differences in predictive validity. See also the significant though narrow differences in the AUCs of OVP and other, correlated, predictors using one-time assessments and fixed followups (Howard & Dixon, 2012).

Results

Table 1 shows that 51% of surviving offenders were reassessed within the 4-month interval recommended in practice guidelines (Ministry of Justice, 2007), 79% within one year and 84% within two years. The proportions with any change in the 62 dynamic risk items were 31%, 58% and 66% at these respective intervals. Reassessments were therefore frequent but far from universal, with a significant minority of reassessments failing to identify any change, confirming the non-ideal nature of assessment practice.

Table 2 shows that four of the seven OVP risk factor scores changed in over 10% of pairs of successive assessments, and one other changed in over 5% of such pairs. Alcohol misuse showed the greatest mean absolute change and the greatest fall in net score; reductions in alcohol misuse score accounted for about half of the total net fall in OVP score. Temper control changed less often, but changes in this score also usually indicated reductions in risk, while accommodation, employability and attitude changes were moderately frequent. Offenders' statuses on the two yes/no questions - recognising the impact of offending and psychiatric treatment - seldom changed.

Changes in total score over the course of supervision were associated with recidivism. Table 3 shows that mean dynamic risk scores fell for all offenders, but fell by more among nonreoffenders than reoffenders even though the latter group commenced community supervision with higher scores and therefore more opportunity for score decreases. Table 4 shows C scores rose when OVP score changes were accounted for. OGRS3 and OVP's 60-point static scale are included as comparators. The C advantage gained from using the total, static/dynamic, OVP score rather than the static OVP score increased when current dynamic scores were used rather than initial-only scores, from 0.0134 to 0.0182 for all violence and 0.0054 to 0.0136 for homicide/wounding, relative increases of 36% and 152%

respectively. Rice and Harris (2005) suggested that an AUC of 0.639 represents a medium effect size and an AUC of 0.714 represents a large effect size. Given the similarity of C and AUC, the C increases of 0.0048 and 0.0082 therefore could be characterized as between one-fifteenth and one-ninth of the improvement of 0.075 needed to move from a medium to large effect size.

Table 5 sets out the results of basic Cox regression models considering the initial score and change in score for violent and homicide/wounding reoffending. In the models in Tables 5, 6 and 7, a zero coefficient for a change in score would indicate that the change is uninformative: that is, that using the initial score alone would produce equal predictive validity to the score at the current assessment. A negative coefficient would show that the initial score had greater predictive validity than the current score. Any statistically significant positive coefficient would show that the change in score added some predictive validity, and equal coefficients for the initial score and change in score would indicate that the current score was the optimum predictor. In Table 5, for all violent reoffending, both scores were predictive, with the initial score and change in score being very similar in predictive value, suggesting that using the current score would optimise predictive validity. For homicide/wounding reoffending, the change in score was more predictive than the initial score; in practical terms, this again points to the value of using the current score.

Tables 6 and 7 show the results of pairs of Cox regression models using the weighted subscales in the OVP 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 OVP’s risk factors could be revised in order to maximise predictive validity. 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. In practice, the rarity of homicide/wounding reoffending made significant results less frequent for this outcome, with standard errors much larger in Table 7 than Table 6.

Most results were consistent between violent and homicide/wounding outcomes. The static OVP score was highly predictive in all models. Recognition of the impact of offending was a consistently poor predictor, considering both initial and change scores. Initial accommodation score was fairly predictive, but changes in accommodation were more predictive, indicating the importance of changes in this risk factor. A similar pattern applied to attitudinal problems. Initial employability scores were more predictive than changes, though both were significant, suggesting that employability ratings may be indicative of a more stable underlying trait. Initial alcohol misuse scores were more predictive of violent than homicide/wounding reoffending, though changes in alcohol score were moderately predictive of both. Psychiatric treatment initial scores were more predictive of homicide/wounding than all violence, though this item's very rare (Table 2) score changes were predictive of all violence. Temper control initial scores were predictive, and score changes were highly predictive, with both effects being greater for homicide/wounding.

The results in Table 2 show the extent and degree of changes in each dynamic risk factor, and Table 6 shows the effect of single-point changes in each dynamic risk factor on violent reoffending. The two sets of results can be combined to generate a summary 'acuteness' metric which is the product of the extent to which change occurs and the effect of each point of change. Table 8 sets out the predictors' weights and the length of the risk factor scales (columns (1) and (2)), the results from the above Tables (columns (3) and (5)) and the necessary calculations to standardise correctly for the length of each risk factor scale (columns (4) and (6) to (9)).

In OVP, accommodation appears to play a greater role in the dynamic prediction of violent reoffending than is allowed for by the risk factor weightings. It accounts for 10% (4 of 40 points) of the dynamic score, yet accounts for 24% of the changes in likelihood of reoffending during the followup.

Alcohol misuse, temper control and attitudes all have dynamic roles in proportion to their shares of the 40-point total dynamic score. Scores on recognising the impact of offending and being in psychiatric treatment, which account for 20% (4 points each) of the 40-point score, have very little value as true dynamic risk factors. Impact scores change quite infrequently, and their changes are entirely nonpredictive. While changes in psychiatric treatment status are reasonably predictive, these occur very infrequently. The ‘acuteness’ of each of the four remaining risk factors is roughly proportionate to their share of the 40-point score.

Discussion

This study aimed to investigate whether causal dynamic risk factors are present in OASys assessments, indicating whether the assessed factors have treatment relevance, and whether there are predictive benefits from repeatedly assessing offenders over their period of community supervision. These results demonstrate continuities with those of other studies which linked repeated measures of social/personal risk factors and reoffending data (Brown et al., 2009; Jones et al., 2010; Schrager & Pacheco, 2011; Quinsey et al., 2006; Olver et al., 2007, Olver & Wong, 2011). As in other studies, changes in risk factors hypothesized to be causal dynamic do occur. These changes are incrementally predictive of reoffending, when added to models comparing only static risk factors and initial measures of the dynamic risk factors. The incremental improvements are small when considered in the context of an individual assessment, but should occur repeatedly among the many NOMS offenders assessed using OASys.

This study’s large sample size allows sufficient statistical power to build on these results. As well as testing the incremental predictive validity of changes in a summary risk score (i.e., OVP), the causal nature of each of its component dynamic risk factors were tested. As in some though not all of the existing studies, ratings were made not by researchers but by field staff in the course of their normal duties, demonstrating that it is possible to measure causal dynamic risk factors reliably enough to be of value in operational settings.

The results were generally positive, showing that OVP's total dynamic score and all but two of its component items met Kraemer et al.'s properties of causal dynamic risk factors. Such risk factors are an important part of the value of OASys, and may function similarly in other integrated risk/need assessment and management systems such as LS/CMI. Predictive scores which include causal dynamic risk factors make it feasible to vary the supervision levels of offenders over time, as changes in score upon assessment review will indicate real changes in both recidivism risk and treatment need. Changes in those five of OVP's seven constituent social/personal risk factors which demonstrated causal dynamic properties can also inform reprioritization of treatment places as time passes. Allocating limited correctional resources on the basis of risk predictor scores, as is done in the English and Welsh probation Tiering system, improves the efficacy with which they are distributed among a given offender caseload. Clinicians can utilise increases and (more often) decreases in OVP scores as indicators that individual offenders have become more or less likely to reoffend violently, and thus vary the intensity of supervision required over the remainder of their sentences. Such changes in risk factors or total predictive scores also offer a helpful intermediate outcome measure for evaluations, which can be observed more quickly than proven recidivism outcomes, and offer insight as to why an intervention designed to reduce recidivism eventually succeed or fails. Moreover, "To be perceived as legitimate by our correctional clients, psychological assessments should be... based at least in part on contemporary, dynamic factors within a person's control... [in order to] recognize (and document) that people can change" (Maruna, 2011, p. 673). These benefits of assessment review must nevertheless be set against the cost of conducting reviews; in England and Wales, such assessment policy decisions will be considered carefully by the managers of increasingly autonomous providers of offender services (Ministry of Justice, 2011b).

Further research is required in this emerging area, whether in England and Wales or in the other

European jurisdictions which use OASys-like assessment tools. This study did not compare the frequency and extent of changes in risk factors occurring during routine supervision (which every offender subject to ongoing OASys assessment receives, to at least some extent) with changes during enhanced activity such as offending behaviour programmes or substance misuse treatment. Indeed, estimates of the effects of correctional activity would ideally also involve assessment of offenders receiving no correctional intervention at all. However, obtaining the information to accurately assess such offenders would be challenging, as contact logs and case notes would not be maintained as they are on supervised offenders. While this study was community-based, and could therefore study associations with reoffending, risk assessment and treatment practice would be further enhanced by knowing whether valid causal changes also occur in custody. Assessments upon custodial entry and discharge could be combined with the probation-based measures of this study to determine whether apparent changes in risk while in custody indicate risk changes upon release. Finally, service providers' decisions on whether and how often to review assessments might be informed by study of the total costs and benefits associated with ongoing assessment activity, akin to the recent economic evaluation of Canadian corrections (Conference Board of Canada, 2009).

Three methodological caveats should be noted. First, given that reassessment did not always happen when it should have, as shown in Table 1, it may be the case that there were systematic differences in quality between assessments which were reviewed and those which were not, affecting the generalizability of the results. While NOMS maintains quality assurance procedures, this issue is likely to continue in the future, and future research could perhaps identify probation offices or geographic areas with particularly good data completion standards in order to test how results change when best practice is followed. Second, OVP, like other predictive scales, was constructed using static and initial-only measures of dynamic risk factors. It is possible that the composition of such scales would have been different if the method used had incorporated time-specific covariates. Finally, this study assumed that increases and decreases in dynamic risk scores had equivalent effects. It is however plausible that assessors' ability to detect improvements in existing problem areas could differ from their ability to detect deteriorations in previously unproblematic areas, and this could be tested by measuring score increases and decreases as separate time-dependent covariates.

References

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

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: LexisNexis / Matthew Bender.

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

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

Beech, A. R., & Ward, T. (2004). The integration of etiology and risk in sexual offenders: A theoretical framework. *Aggression and Violent Behavior, 10*, 31-63. doi: 10.1016/j.avb.2003.08.002

Brown, S. L., St. Amand, M. D., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: A three-wave prospective study. *Law & Human Behavior, 33*, 25-45. doi: 10.1007/s10979-008-9139-7

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36*, 567-590. doi: 10.1177/0093854809333610

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

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

Gendreau, P., Smith, P., & Thériault, Y. L. (2009). Chaos theory and correctional treatment: Common sense, correctional quackery, and the law of fartcatchers. *Journal of Contemporary Criminal Justice, 25*, 384-396. doi: 10.1177/1043986209344552

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

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

Hanson, R. K., & Howard, P. D. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior, 34,* 275-281. doi: 10.1007/s10979-010-9227-3

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

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

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

387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413-425. doi: 10.1037/1040-3590.15.3.413

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

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

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

Howard, P. (2006). *The Offender Assessment System: An evaluation of the second pilot* (Home Office Research Findings 278). Retrieved from http://www.homeoffice.gov.uk/rds/pdfs06/r278.pdf

Howard, P. (2009). *Improving the prediction of reoffending using the Offender Assessment System* (Ministry of Justice Research Summary 2/09). Retrieved from webarchive.nationalarchives.gov.uk/20110201125714/http://www.justice.gov.uk/publications/docs/oasys-research-summary-02-09.pdf

Howard, P. D., & Dixon, L. (2011). Developing an empirical classification of violent offenses for use in the prediction of recidivism in England and Wales. *Journal of Aggression, Conflict and Peace Research, 3*, 141-154. doi: 10.1108/17596591111154176

Howard, P. D., & Dixon, L. (2012). The construction and validation of the OASys Violence Predictor: Advancing violence risk assessment in the English and Welsh correctional services. *Criminal Justice and Behavior, 39,* 287-307. doi: 10.1177/0093854811431239

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

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

Jones, N. J., Brown, S. L, & Zamble, E. (2010). Predicting criminal recidivism in adult male offenders: Researcher versus parole officer assessment of dynamic risk. *Criminal Justice and Behavior, 37*, 860-882. doi: 10.1177/0093854810368924

Kattan, M. W. (2003). Judging new markers by their ability to improve predictive accuracy. *Journal of the National Cancer Institute, 95*, 634-635. doi: 10.1093/jnci/95.9.634

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*, 337-343. doi: not available.

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

Maruna, S. (2011). Why do they hate us? Making peace between prisoners and psychology. *International Journal of Offender Therapy and Comparative Criminology, 55*, 671-675. doi: 10.1177/0306624X11414401

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

Ministry of Justice (2007). *National Standards for the management of offenders: Standards and implementation guidance.* Retrieved from http://webarchive.nationalarchives.gov.uk/20100105030830/http://www.probation.homeoffice.gov.uk/files/pdf/OM%20National%20Standards%202007.pdf

Ministry of Justice (2009). *Offender caseload management statistics 2008*. Retrieved from http://www.justice.gov.uk/statistics/prisons-and-probation/omcs-annual

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

Monahan, J., Steadman, H., Robbins, P., Appelbaum, P., Banks, S., Grisso, T., Heilbrun, K., Mulvey, E., Roth, L., & Silver, E. (2005). An actuarial model of violence risk assessment for persons with mental disorders. *Psychiatric Services, 56*, 810-815. doi: not available.

Moore, R. (2009). *The internal reliability and construct validity of the Offender Assessment System (OASys)*

(Ministry of Justice Research Summary 6/09). Retrieved from webarchive.nationalarchives.gov.uk/20110201125714/http://www.justice.gov.uk/publications/docs/oasys-research-summary-06-09.pdf

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

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

National Probation Service (2008). *National rules for tiering cases and associated guidance*. (Probation Circular 08/08.) Retrieved from webarchive.nationalarchives.gov.uk/20100105030830/http://www.probation.homeoffice.gov.uk/files/pdf/PC08%202008.pdf

Olver, M. E., Wong, S. C. P., Nicolaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale – Sexual Offender version. Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318-329. doi: 10.1037/1040-3590.19.3.318

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

Palmer, E. J, McGuire, J., Hatcher, R. M., Hounsome, J. C., Bilby, C. A. L., & Hollin, C. R. (2009). Allocation to offending behavior programs in the English and Welsh probation service. *Criminal Justice and Behavior, 36*, 909-922. doi: 10.1177/0093854809338889

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

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.

Quinsey, V. L., Jones, G. B., Book, A. S., & Barr, K. N. (2006). The dynamic prediction of antisocial behavior among forensic psychiatric patients. A prospective field study. *Journal of Interpersonal Violence, 21*, 1539-1565. doi: 10.1177/0886260506294238

Raykar, V. C, Steck, H., Krishnapuram, B., Dehing-Oberije, C., & Lambin, P. (2008). On ranking in survival analysis: Bounds on the Concordance Index. *Advances in Neural Information Processing Systems, 20*, 1209-1216. doi: not available.

Schlager, M. D., & Pacheco, D. (2011). An examination of changes in LSI-R scores over time: Making the case for needs-based case management. *Criminal Justice and Behavior,38,* 541-553. doi: 10.1177/0093854811402300

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

Taxman, F. S., & Thanner, M. (2006). Risk, need and responsivity (RNR): It all depends. *Crime and Delinquency, 52*, 28-51. doi: 10.1177/0011128705281754

Thornton, D. (2007). *Scoring Guide for Risk Matrix 2000.9/SVC.* Retrieved from http://www.birmingham.ac.uk/Documents/college-les/psych/RM2000scoringinstructions.pdf

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

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

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

Wong, S. C. P., & Gordon, A. E. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy and the Law, 12*, 279-309. doi: 10.1037/1076-8971.12.3.279

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767. doi 10.1037/a0020473

Table 1

Life table tracing violent reoffending, censoring, reassessment on OASys and dynamic risk factor change over a 5-year followup

Time period	Number at start of period	Number censored	Number of non-censored reoffending (% hazard)	Cumulative % with no reassessment by end of period	Cumulative % with no change in dynamic RFs by end of this period
0-4 months	206,214	8,294	16,819 (8.5%)	48.6%	68.9%
4-8 months	181,101	6,597	11,132 (6.4%)	27.4%	50.7%
8-12 months	163,372	4,266	8,095 (5.1%)	21.5%	42.0%
12-16 months	151,101	6,936	6,651 (4.6%)	18.7%	38.0%
17-20 months	137,424	19,591	4,688 (4.0%)	17.1%	35.6%
20-24 months	113,145	18,801	3,489 (3.7%)	16.1%	34.1%
24-28 months	90,855	15,604	2,450 (3.3%)	15.4%	33.2%
28-32 months	72,801	13,205	1,808 (3.0%)	14.9%	32.5%
32-36 months	57,788	11,705	1,164 (2.5%)	14.4%	32.0%
36-40 months	44,919	9,651	844 (2.4%)	14.0%	31.6%
40-44 months	34,424	8,980	595 (2.3%)	13.7%	31.3%
44-48 months	24,849	7,749	346 (2.0%)	13.4%	31.0%

Note. Dynamic RFs = all 62 OASys dynamic risk factor items, covering accommodation, education training and employability, financial management and income, relationships, lifestyle and associates, drug misuse, alcohol misuse, emotional well-being, thinking and behaviour, and attitudes domains. Change refers to any change in at least one of these items. 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”.

Table 2

Changes in OVP 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
Total score (100)	*39.7 (13.8)*	*1.28 (3%)*	*-0.43 (-1%)*	*39.4%*
Static score (60)	27.7 (9.2)	n/a	n/a	n/a
Total dynamic score (40)	*11.9 (7.2)*	*1.28 (11%)*	*-0.43 (-4%)*	*39.4%*
Recognises impact of offending on victim/community/society (4)	0.85 (1.64)	0.07 (8%)	-0.01 (-1%)	1.7%
Accommodation (4)	1.11 (1.46)	0.28 (25%)	-0.05 (-4%)	13.9%
Employability (6)	2.76 (2.00)	0.23 (8%)	-0.06 (-2%)	13.8%
Alcohol misuse (10)	3.33 (3.85)	0.46 (14%)	-0.21 (-6%)	11.8%
Psychiatric treatment current/pending (4)	0.21 (0.89)	0.02 (10%)	0.00 (2%)	0.5%
Temper control (6)	2.09 (2.29)	0.20 (10%)	-0.09 (-4%)	6.3%
Attitudes (6)	1.57 (1.33)	0.20 (13%)	-0.02 (-1%)	15.4%

Note. Initial assessment N = 196,493. Change assessment N, and N of successive assessment pairs = 378,596. All mean net changes were significantly different from zero at p<.0001. 'Any change' involves any change in the items comprising the risk factor and/or total score, including offsetting changes in multiple items. Time periods between assessments ranged from 1 to 1,691 days, with mean 131.68 days, SD 125.83 days.

Table 3

Initial scores and changes in score by final assessment, for reoffenders and nonreoffenders

Reoffence type and predictor score used, and followup type and outcome	No. of cases	Initial static/dynamic score (100-point maximum)			Initial dynamic score (40-point maximum)			Change from initial to final score		
		Mean	SD	SE	Mean	SD	SE	Mean *(% of mean initial)*	SD	SE
Violent reoffending										
Followups involving one assessment only										
No reoffending	28,468	38.7	14.1	0.08	11.32	7.43	0.04	n/a	n/a	n/a
Reoffending	21,270	48.5	12.3	0.08	15.30	7.59	0.05	n/a	n/a	n/a
Followups involving multiple assessments										
No reoffending	116,101	36.6	13.2	0.04	10.87	6.81	0.02	-1.23 *(-3.4%)*	4.15	0.01
Reoffending	30,654	46.0	11.7	0.07	14.13	7.22	0.04	-0.71 *(-1.5%)*	4.12	0.02
Homicide and wounding reoffending										
Followups involving one assessment only										
No reoffending	37,310	41.0	14.4	0.07	12.20	7.68	0.04	n/a	n/a	n/a
Reoffending	524	50.1	12.1	0.53	15.58	7.73	0.34	n/a	n/a	n/a
Followups involving multiple assessments										
No reoffending	157,851	39.3	13.6	0.03	11.83	7.12	0.02	-0.86 *(-2.2%)*	4.53	0.01
Reoffending	808	48.8	11.8	0.41	15.05	7.56	0.27	-0.28 *(-0.6%)*	4.23	0.15

Table 4

Predictive validity of OGRS3 and initial and current OVP scores

Reoffending outcome	Concordance Index by predictor and scoring method			
	OGRS3	OVP		
		Static-only	Initial	Current
Violent	0.6825	0.7048	0.7182	0.7230
Homicide & wounding	0.6862	0.7313	0.7367	0.7449

Note. *N* = 196,493 offenders.

Table 5

Cox regression models: total OVP score as a predictor of violent and homicide/wounding reoffending outcomes

Reoffending outcome	Parameter estimates					
	Initial score			Change in score		
	Beta	SE	Hazard ratio	Beta	SE	Hazard ratio
Violent	0.0597	.0003	1.062	0.0563	.0016	1.058
Homicide & wounding	0.0670	.0021	1.069	0.0793	.0093	1.083

Note. Beta = effect size per point of predictor. SE = standard error of Beta. Hazard ratio = ratio of hazards for scores *x+1* and *x*. *N* = 575,089 assessments (violent reoffending), 635,697 assessments (homicide/wounding reoffending).

Table 6

Cox regression models: risk factors in OVP 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	94.3	0.075	***	0.001	1.078	90.0
Dynamic: initial assessment	Recognises impact (4)	-0.001		0.003	0.995	0.98	-0.007	*	0.003	0.993	0.97
	Accommodation (4)	0.040	***	0.003	1.041	1.17	0.051	***	0.001	1.052	1.23
	Employability (6)	0.057	***	0.003	1.059	1.41	0.056	***	0.005	1.058	1.40
	Alcohol misuse (10)	0.047	***	0.001	1.048	1.60	0.051	***	0.002	1.053	1.65
	Psychiatric treatment (4)	0.020	***	0.005	1.020	1.09	0.019	***	0.004	1.020	1.09
	Temper control (6)	0.032	***	0.002	1.033	1.22	0.039	***	0.002	1.040	1.27
	Attitudes (6)	0.045	***	0.004	1.046	1.31	0.048	***	0.006	1.049	1.32
Dynamic: change from initial to most recent assessment	Recognises impact (4)						-0.003		0.010	1.001	1.00
	Accommodation (4)						0.065	***	0.006	1.067	1.30
	Employability (6)						0.035	***	0.007	1.035	1.15
	Alcohol misuse (10)						0.047	***	0.003	1.048	1.59
	Psychiatric treatment (4)						0.047	**	0.015	1.048	1.21
	Temper control (6)						0.062	***	0.005	1.063	1.44
	Attitudes (6)						0.066	***	0.008	1.069	1.49

Note. Recognises impact and psychiatric treatment can only be scored 0 or 4. Alcohol misuse can only be scored 0, 3, 5, 8 or 10. Temper control can only be scored 0, 3 or 6.

* $p < .05$. ** $p < .01$. *** $p < .001$.

$N = 575{,}089$ assessments.

Table 7

Cox regression models: risk factors in OVP as predictors of homicide/wounding 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)	0.092	***	0.004	1.097	254.2	0.090	***	0.004	1.095	226.8
Dynamic: initial assessment	Recognises impact (4)	-0.016		0.017	0.984	0.94	-0.014		0.017	0.986	0.95
	Accommodation (4)	0.039	*	0.019	1.040	1.17	0.049	*	0.021	1.050	1.22
	Employability (6)	0.072	***	0.016	1.075	1.54	0.071	***	0.016	1.074	1.53
	Alcohol misuse (10)	0.011		0.007	1.011	1.12	0.015	*	0.008	1.015	1.16
	Psychiatric treatment (4)	0.067	*	0.027	1.069	1.28	0.061	*	0.027	1.062	1.26
	Temper control (6)	0.077	***	0.013	1.080	1.59	0.094	***	0.014	1.099	1.76
	Attitudes (6)	0.034		0.023	1.035	1.23	0.037		0.024	1.038	1.25
Dynamic: change from initial to most recent assessment	Recognises impact (4)						0.103	*	0.051	1.108	1.51
	Accommodation (4)						0.065	*	0.033	1.068	1.38
	Employability (6)						0.042		0.042	1.043	1.34
	Alcohol misuse (10)						0.038	*	0.018	1.039	1.47
	Psychiatric treatment (4)						-0.041		0.082	0.959	0.85
	Temper control (6)						0.149	***	0.032	1.160	2.44
	Attitudes (6)						0.084		0.046	1.088	1.66

Note. Recognises impact and psychiatric treatment can only be scored 0 or 4. Alcohol misuse can only be scored 0, 3, 5, 8 or 10. Temper control can only be scored 0, 3 or 6.

* $p < .05$. ** $p < .01$. *** $p < .001$.

N = 635,697 assessments.

Table 8

Acuteness of dynamic risk factors in OVP as a predictor of violent reoffending

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)
Impact	4	2	-.003	-.006	.07	.018	-.00011	-.0002	<1
Accommodation	4	8	.065	.033	.28	.070	.00231	.0185	24
Employability	6	8	.035	.026	.23	.038	.00099	.0079	10
Alcohol misuse	10	4	.047	.118	.46	.046	.00543	.0217	28
Psychiatric treatment	4	2	.047	.095	.02	.005	.00048	.0010	1
Temper control	6	2	.062	.186	.20	.033	.00614	.0123	16
Attitudes	6	8	.066	.050	.20	.038	.00188	.0150	20
Total	*40*							*.0762*	*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 Table 6. Weighted mean absolute changes (5) are from Table 2. (9) = (8) / sum of all (8) values.