Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool

Seena Fazel, Zheng Chang, Thomas Fanshawe, Niklas Långström, Paul Lichtenstein, Henrik Larsson, Susan Mallett

Summary

Background More than 30 million people are released from prison worldwide every year, who include a group at high risk of perpetrating interpersonal violence. Because there is considerable inconsistency and inefficiency in identifying those who would benefit from interventions to reduce this risk, we developed and validated a clinical prediction rule to determine the risk of violent offending in released prisoners.

Methods We did a cohort study of a population of released prisoners in Sweden. Through linkage of population-based registers, we developed predictive models for violent reoffending for the cohort. First, we developed a derivation model to determine the strength of prespecified, routinely obtained criminal history, sociodemographic, and clinical risk factors using multivariable Cox proportional hazard regression, and then tested them in an external validation. We measured discrimination and calibration for prediction of our primary outcome of violent reoffending at 1 and 2 years using cutoffs of 10% for 1-year risk and 20% for 2-year risk.

Findings We identified a cohort of 47 326 prisoners released in Sweden between 2001 and 2009, with 11 263 incidents of violent reoffending during this period. We developed a 14-item derivation model to predict violent reoffending and tested it in an external validation (assigning 37 100 individuals to the derivation sample and 10 226 to the validation sample). The model showed good measures of discrimination (Harrell’s c-index 0·74) and calibration. For risk of violent reoffending at 1 year, sensitivity was 76% (95% CI 73–79) and specificity was 61% (95% CI 60–62). Positive and negative predictive values were 21% (95% CI 19–22) and 95% (95% CI 94–96), respectively. At 2 years, sensitivity was 67% (95% CI 64–69) and specificity was 70% (95% CI 69–72). Positive and negative predictive values were 37% (95% CI 35–39) and 89% (95% CI 88–90), respectively. Of individuals with a predicted risk of violent reoffending of 50% or more, 88% had drug and alcohol use disorders. We used the model to generate a simple, web-based, risk calculator (OxRec) that is free to use.

Interpretation We have developed a prediction model in a Swedish prison population that can assist with decision making on release by identifying those who are at low risk of future violent offending, and those at high risk of violent reoffending who might benefit from drug and alcohol treatment. Further assessments in other populations and countries are needed.

Funding Wellcome Trust, the Swedish Research Council, and the Swedish Research Council for Health, Working Life and Welfare.

Copyright © Fazel et al. Open Access article distributed under the terms of CC BY.

Introduction

To reduce the mortality and morbidity burden associated with interpersonal violence, the identification of, and intervention with, prisoners at high risk of perpetrating violence provides an approach with considerable public health and safety benefits.1 Repeat offending rates remain high in many high-income countries,2 and have not followed the downward trend of violence reported in these countries.1 In England and Wales, prisoners have been reconvicted at a 2-year rate of 55%–60% for the past decade.1 With about 30 million individuals entering and leaving prison per year worldwide,1 the contribution of this population to societal violence is high, and an estimated 20% of all arrests in the USA,2 and 18% of new crimes in the UK,3 are by former prisoners.

To identify individuals who are at the highest risk of reoffending and most in need of interventions to reduce future criminality, criminal justice agencies in most high-income and middle-income countries have used actuarial and clinically informed decision aids. These aids assist with decisions about sentencing, entry into specific programmes for prison treatment and aftercare, and the timing of release from detention and need for supervision on release. More than 300 of these risk assessment tools exist, but they are limited by low to moderate accuracy,7 financial and non-financial competing interests affecting the research evidence,9 and inconsistent definitions of risk classifications.1 Typically, these tools identify prisoners at low, medium, and high risk of repeat offending on the basis of an assessment weighted towards historical non-modifiable factors. Many of the tools are expensive to use and require training to administer. Another problem is that they are usually developed without predetermined protocols, the use of which enhances transparency and reduces bias by clarifying key elements in research design before data acquisition or analysis, and hence increasing
Research in context

Evidence before this study
We searched PubMed for articles published in the previous 5 years (from Jan 1, 2010, to July 27, 2015) for the terms violence, risk assessment, and review in MeSH terms and all fields based on all languages: (“violence”[MeSH Terms] OR “violence”[All Fields]) AND (“risk assessment”[MeSH Terms]) OR (“risk”[All Fields] AND “assessment”[All Fields]) OR (“risk assessment”[All Fields]) AND (Review[ptyp]). We identified six systematic reviews, two reporting on methods-related limitations of risk assessment, two focusing on selected populations, and two comparing the performance of tools. One review highlighted the large variation between tools, with regard to what proportion of prisoners who are categorised as high risk actually reoffend. A further review summarised the predictive accuracy of the most commonly used tools for assessment of risk of violence, and showed low to moderate accuracy on a range of performance metrics.

Added value of this study
We have developed a risk score for violent reoffending in a total population of released prisoners that is externally validated and includes modifiable risk factors. The novel features of this study are that it uses the methods to develop the risk score on the basis of TRIPOD guidelines, and that it is a brief, easy to use, and scalable tool. Additionally, for the first time for violent reoffending, this tool has been translated into a freely available web calculator.

Implications of all the available evidence
Criminal justice and community health services can potentially improve reoffending outcomes, particularly if they work together on modifiable risk factors, and this risk score can assist in identifying prisoners on release who are at high risk of violent reoffending, and might benefit from interventions to reduce crime and treat substance use disorders.

the quality of prognosis research. A further limitation of these risk assessment tools is that they do not include information about psychiatric disorders and substance misuse in a consistent way—many do not include such conditions, and others use varying definitions. Because individuals with psychiatric disorders constitute up to 50% of the worldwide prison population, experience a large treatment gap, and are associated with reoffending, their inclusion in tools could improve accuracy and lead to targeting those who would benefit the most from interventions. At the same time, scalable tools are needed because current approaches typically require a face-to-face assessment and take hours to complete. In England and Wales, for example, the officially sanctioned tool used to assess such prisoners includes at least 76 items. In this study, we report the development and validation of a clinical prediction rule to determine the risk of violent offending in released prisoners.

Methods

Study design and participants
We did a cohort study of a large population of released prisoners in Sweden, identified via the Swedish Prison and Probation Service. We followed up each individual from the day of their release until first violent reoffending, reincarceration, death, emigration, or end of the study. The study was approved by the Regional Ethics Committee at Karolinska Institutet, Sweden (2009/939-31/5).

Measurement of predictors
Individuals within the study cohort were linked to several national population-based registers to obtain information on risk factors, with unique personal identification numbers enabling accurate linkage (appendix p 1). On the basis of existing evidence that groups risk factors into criminal history, sociodemographic, and clinical domains, we decided a priori to consider risk factors in three groups of decreasing levels of priority. Additionally, we categorised length of incarceration, highest education, and disposable income into categorical variables (table 1), and clarified these in a predetermined protocol drawn up by the investigators before statistical analyses (appendix p 14–17). Risk factors were designated a group number 1, 2, or 3, referring to the strength of previous evidence supporting their inclusion in a prediction score, where 1 was the highest strength. The overall degree of socioeconomic deprivation was given a standardised, normalised score (including rates of welfare recipiency, unemployment, poor education, crime rates, and median income) in an individual’s residential area, with high scores indicating high levels of deprivation.

We identified lifetime diagnoses of psychiatric disorders before prison release (ie, before and during incarceration) from the Swedish National Patient Register, which provides diagnoses for all inpatient psychiatric hospital admissions in Sweden since 1973 and outpatient care since 2001, according to the International Classification of Diseases (ICD) 8th Revision (ICD-8), 1973–86; 9th Revision (ICD-9), 1987–96; or 10th Revision (ICD-10), 1997–2009. We included personality disorders in the mental health disorder category. We recorded diagnoses of psychiatric disorders as five binary variables (present or absent) based on any lifetime diagnosis before prison release (for ICD codes, see appendix p 1). Comorbidity was included, so if a participant had a comorbid drug or alcohol use disorder, this was coded as present even if the prisoner had an Axis I diagnosis.
Measurement of outcomes
Our primary outcome was first occurrence of violent reoffending at 1 and 2 years, defined as the first new conviction for any violent crime after release from prison through the National Crime Register. We used conviction data because the Swedish criminal code finds individuals as guilty regardless of the presence of any mental health disorder, although sentencing might be informed by such disorders. In keeping with previous studies, violent crime was defined as homicide, assault, robbery, arson, any sexual offence (rape, sexual coercion, child molestation, indecent exposure, or sexual harassment), illegal threats, or intimidation. If the date of the crime was not recorded, the date of the conviction was used. Reoffending for any crime (violent and non-violent) was a specified secondary outcome.

Statistical analysis
Our analyses involved a two-stage strategy: derivation of the models followed by their validation. Before these analyses, we selected a subsample on the basis of the residential location of the individual at the year of imprisonment to act as the external validation sample selected on a geographical basis (appendix pp 1–2). The remaining data were used to derive models that could be used to predict each of the outcomes (the derivation sample).

We used multivariable Cox proportional hazard regression to investigate the association of measured risk factors and violent reoffending and account for different follow-up times. On the basis of the risk factors identified above, we used a three-step approach to develop the most parsimonious model while reducing variable selection within it, and one that showed face validity and allowed for the inclusion of additional risk factors associated with outcomes (appendix p 2).

Multiple imputation was used to replace missing values for risk factor variables, with regression models that used all other risk factors, the main outcome, and the Nelson–Aalen cumulative hazard function. Any missing values were assumed to be missing at random in the multiple imputation model (ie, associated only with other measured variables). We did 20 imputations, and computed estimates of coefficients by combining information across all imputed datasets at each stage of the variable selection process and in the final model.

Once the final models were identified using the variable selection procedure above, we assessed the predictive ability of the model in terms of discrimination and calibration. First, we used Harrell’s c-index as an overall measure of discrimination, which refers to the ability of the risk prediction model to differentiate between individuals who do and do not experience the outcome. The c-index varies from 0·5 to 1·0, with 1·0 representing perfect discrimination. For prespecified timepoints we additionally calculated the statistics for the area under the receiver operating characteristics curve (AUC). Then we took the regression coefficients for each variable from the final models and used these as predicted hazard ratios, which we combined with the baseline survivor function for each outcome at 1 year and 24 months. We then computed estimates of coefficients by combining the multiple imputation model (ie, associated only with other measured variables). We did 20 imputations, and repeated the entire set of analyses. We did all analyses using Stata (version 14.2).

Table 1: Baseline characteristics of the study population and group of risk factors

<table>
<thead>
<tr>
<th>Risk factor variable</th>
<th>Derivation sample (n=37 100)</th>
<th>Validation sample (n=10 226)</th>
<th>Risk factor group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>34 356 (93%)</td>
<td>9484 (93%)</td>
<td>1</td>
</tr>
<tr>
<td>Female</td>
<td>2597 (7%)</td>
<td>716 (7%)</td>
<td>1</td>
</tr>
<tr>
<td>Age (years)</td>
<td>36 (27–46)</td>
<td>36 (26–46)</td>
<td>1</td>
</tr>
<tr>
<td>Immigrant status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11 410 (21%)</td>
<td>3106 (30%)</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>25 690 (69%)</td>
<td>7120 (70%)</td>
<td>1</td>
</tr>
<tr>
<td>Length of incarceration (months)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6</td>
<td>25 584 (69%)</td>
<td>7179 (70%)</td>
<td>1</td>
</tr>
<tr>
<td>6–11</td>
<td>6086 (16%)</td>
<td>1690 (17%)</td>
<td>1</td>
</tr>
<tr>
<td>12–23</td>
<td>3771 (10%)</td>
<td>920 (9%)</td>
<td>1</td>
</tr>
<tr>
<td>≥24</td>
<td>1659 (4%)</td>
<td>457 (4%)</td>
<td>1</td>
</tr>
<tr>
<td>Violent index offence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14 182 (38%)</td>
<td>3755 (37%)</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>22 918 (62%)</td>
<td>6471 (63%)</td>
<td>1</td>
</tr>
<tr>
<td>Previous violent crime</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>19 714 (53%)</td>
<td>5358 (52%)</td>
<td>1</td>
</tr>
<tr>
<td>No</td>
<td>17 386 (47%)</td>
<td>4868 (48%)</td>
<td>1</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>22 565 (65%)</td>
<td>6134 (64%)</td>
<td>2</td>
</tr>
<tr>
<td>Married</td>
<td>14 535 (35%)</td>
<td>4092 (36%)</td>
<td>2</td>
</tr>
<tr>
<td>Education (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;9</td>
<td>16 697 (48%)</td>
<td>4847 (50%)</td>
<td>2</td>
</tr>
<tr>
<td>9–11</td>
<td>25 816 (66%)</td>
<td>4203 (44%)</td>
<td>2</td>
</tr>
<tr>
<td>≥12</td>
<td>2111 (6%)</td>
<td>538 (6%)</td>
<td>2</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8781 (25%)</td>
<td>2416 (25%)</td>
<td>2</td>
</tr>
<tr>
<td>No</td>
<td>28 319 (75%)</td>
<td>7810 (75%)</td>
<td>2</td>
</tr>
<tr>
<td>Disposable income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>218 (1%)</td>
<td>59 (1%)</td>
<td>2</td>
</tr>
<tr>
<td>Zero</td>
<td>1942 (6%)</td>
<td>442 (5%)</td>
<td>2</td>
</tr>
<tr>
<td>Low (&lt;20th percentile)</td>
<td>18 189 (53%)</td>
<td>5130 (53%)</td>
<td>2</td>
</tr>
<tr>
<td>Medium (20th–80th percentile)</td>
<td>13 859 (40%)</td>
<td>3857 (40%)</td>
<td>2</td>
</tr>
<tr>
<td>High (&gt;80th percentile)</td>
<td>406 (1%)</td>
<td>100 (1%)</td>
<td>2</td>
</tr>
<tr>
<td>Neighbourhood deprivation (standardised score)</td>
<td>0.39 (–1.18 to 1.47)</td>
<td>0.35 (–1.16 to 1.61)</td>
<td>2</td>
</tr>
<tr>
<td>Mental health disorders*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol use disorder</td>
<td>8094 (22%)</td>
<td>2150 (21%)</td>
<td>3</td>
</tr>
<tr>
<td>Drug use disorder</td>
<td>8680 (23%)</td>
<td>2355 (23%)</td>
<td>3</td>
</tr>
<tr>
<td>Any mental health disorder</td>
<td>8281 (22%)</td>
<td>2658 (26%)</td>
<td>3</td>
</tr>
<tr>
<td>Any severe mental health disorder</td>
<td>1288 (3%)</td>
<td>330 (3%)</td>
<td>3</td>
</tr>
<tr>
<td>Any developmental or childhood disorder</td>
<td>1179 (3%)</td>
<td>344 (3%)</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes: *Refers to most recent ICD-10 clinical diagnoses (including comorbidity) in the previous 5 years (including in prison). 2476 (7%) individuals in the derivation sample and 638 (6%) in the validation sample had missing values for civil status, highest education, employment, disposable income, and neighbourhood deprivation. The number of missing values were the same for all these variables, because they were from the same register.
2 years after release to derive the corresponding absolute predicted probabilities. The equation used for the model is reported in the appendix (p 8).

We calculated sensitivity, specificity, positive, and negative predictive values separately with prespecified binary thresholds. Predicted probabilities of 10% and 20% were used for violent reoffending within 1 and 2 years, respectively, and 50% was used for any reoffending within 1 and 2 years after release. Individuals who were censored before the end of the follow-up period (either 1 year or 2 years) were excluded from these calculations. For internal validation, we used bootstrapping methods (200 times) to provide estimates corrected for model performance for population sampling, and to quantify potential for overfitting. \(^\text{22,23}\) The predicted probability also allowed us to stratify individuals into three pre-specified risk groups: low (<10%), medium (10–50%), and high (>50%) according to the individual’s predicted risk of violent reoffending within 2 years. We chose these values on the basis of a review\(^\text{9}\) that reported that the median annual rate of violence in high-risk individuals identified by risk assessment instruments was 13% (IQR 7%–19%).

We assessed calibration (which indicates whether predicted risks agree with observed risks) by plotting the predicted observed risk of outcome versus the observed risk of outcome. We also calculated Brier scores, defined as the average quadratic difference between the predicted probability and the binary outcome. \(^\text{24}\) The Brier score ranges from 0 to 1, with lower scores indicating better calibration. To help assess model performance, we also calculated Brier scores in two other scenarios: 1) assigning zero predicted probability to each individual, and 2) assigning the mean predicted probability across the whole cohort to each individual.

We applied the coefficients and baseline survival function of the final model developed from the derivation sample to the validation sample. All variables were rounded to four decimal places (appendix p 8). We then calculated predicted probabilities for each individual, and assessed the discrimination and calibration for violent reoffending within 1 and 2 years. We used STATA (version 12) for all analyses and followed the TRIPOD statement. \(^\text{25}\)

**Role of the funding source**

The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. ZC had full access to all the data in the study and, with SF, had final responsibility for the integrity of the data, the accuracy of the data analysis, and the decision to submit for final publication.

**Results**

We identified a cohort of 47 326 prisoners who were imprisoned since Jan 1, 2000, and released before Dec 31, 2009, with 11 263 incidents of violent reoffending during the study period. Of the total cohort, we assigned 37 100 participants to the derivation sample and 10 226 to the external validation sample. Baseline characteristics of individuals in both samples were similar (table 1). 2476 individuals in the derivation sample and 638 in the validation sample had missing values for the categories of marital status, highest education, employment, disposable income, and neighbourhood deprivation. In the derivation sample, 8883 (24%) reoffended for a violent crime and 21 739 (59%) reoffended for any crime during the mean follow-up of 3·2 years (SD 2·6). The corresponding figures in the external validation samples were 2380 (23%) and 5927 (58%), respectively (for risk factors and age distribution, see appendix pp 3, 4, 10). The estimated probability of violent reoffending at 1, 2, and 5 years after release was 11%, 18%, and 31% respectively (figure 1). Overall, more than 70% of the repeat violent offenses were for assault or robbery, with 2% for sexual offences and 1% for homicide (appendix p 5).

Risk factors included in the final model were male sex, younger age, non-immigrant status, shorter length of incarceration, violent index (or most recent) offence, previous violent crime, being never married, fewer years of formal education, being unemployed before prison, low disposable income, living in an area of higher neighbourhood deprivation, and diagnoses of alcohol use disorder, drug use disorder, any mental disorder, and any severe mental disorder (appendix p 6). Younger age and higher neighbourhood deprivation were associated with increased predicted risks of violent reoffending at 1 year and 2 years (appendix pp 10–11). The model showed good overall discrimination (Harrell’s c-index=0·74), and performed well on measures of discrimination for violent reoffending within 1 and 2 years (figure 2).

For violent reoffending within 1 year after prison release, the external validation model showed good discrimination with an AUC of 0·75 (figure 2) and good calibration (figure 3), except in a small group of

---

**Figure 1: Estimated probability of violent reoffending in released prisoners in Sweden**

Survivor function

95% CI
individuals (n=18, <1%) for whom the model overestimated the risk at 50% at 1 year. The Brier score (0.095) was lower than that using the mean predicted probability (0.106) or using zero (0.120). For the risk of violent reoffending at 1 year after prison release, sensitivity was 76% (95% CI 73–79) and specificity was 61% (60–62). Positive and negative predictive values were 21% (19–22) and 95% (94–96), respectively (table 2).

For the risk of violent reoffending within 2 years after release, the validation model showed good discrimination with an AUC of 0.76 (figure 2), and good calibration (figure 3). The Brier score (0.108) was lower than that with the mean predicted probability (0.126) or with zero (0.141). At 2 years, sensitivity was 67% (95% CI 64–69) and specificity was 70% (69–72). Positive and negative predictive values were 37% (95% CI 33–39) and 89% (88–90), respectively (table 2). The proportion of released prisoners who violently reoffended according to low, medium, and high risk groups was similar in the derivation and validation samples (figure 4). Of those in the highest-risk group (≥50% risk threshold), 935 (88%) of 1066 had a diagnosis of a drug or alcohol use disorder.
compared with 9459 (42%) of 22514 in the medium-risk group (10% to <50%), and 2728 (20%) of 13520 in the low-risk group (<10%).

We did the same analyses for any reoffending as a secondary outcome for risk factors (appendix p 7). The model showed good discrimination and calibration (appendix pp 12, 13) in internal and external validation.

We applied the coefficients to develop an online calculator for predicting risk of violent reoffending called OxRec (Oxford Risk of Recidivism tool; see appendix p 8 for all variables used). This calculator provides both a risk classification (low, medium, or high) and a probability of violent reoffending in 1 and 2 years after prison release. If values are missing, the calculator reports the upper and lower range of estimates of risk allowing for the missing variables.

**Discussion**

We have developed a prediction model and web calculator (OxRec) for the risk of violent reoffending in prisoners on release in Sweden, with good measures of discrimination and calibration. The prediction score uses routinely obtained information, and can assist in decision making when identifying those who could be targeted for crime reduction and substance misuse interventions on release.

Most risk assessment tools that predict reoffending in prisoners (including those with mental health disorders) require several hours to administer, and many of them require training. Some include modifiable psychiatric factors but most do not. Furthermore, many are used in a probabilistic way to determine risk of recidivism despite uncertainties in their validity and precision, and those that provide risk classifications are highly variable in how this translates into rates of reoffending.

The performance statistics of this risk score were not worse than those reported for other tools used in criminal justice and mental health. For 2-year violent reoffending outcomes, the model had a sensitivity of 67% and specificity of 70% in the external validation sample using prespecified risk thresholds. The positive and negative predictive values were 37% and 89%, respectively. The overall c-index was 0.74, and over 2 years, the AUC was 0.76. For the nine most commonly used tools for assessing violence risk in criminal justice and forensic mental health, a 2012 systematic review reported the median AUC of 0.72 and positive predictive value of 41%. This review showed that the specificity for commonly used risk instruments was 36%, lower than that reported here (which was 67%), whereas sensitivity was higher at 91% (compared with 70% for this risk score), but this difference is probably due to different thresholds for risk categories. For example, if we altered our lowest risk threshold from 10% to 6%, then specificity would be 41% and sensitivity 91%. In relation to prediction scores in other areas of medicine, our model also performs similarly. A review of commonly used cardiovascular risk scores reports typical AUCs in the range of 0.70 to 0.75. We opted to present the 2×2 tables showing the
range of true and false positives and negatives. These are easier to interpret than AUCs, which are problematic on their own and mask the different consequences of false negatives and positives.\textsuperscript{27} Furthermore, AUCs are insensitive to changes in model performance, and should not be used to compare different tools.\textsuperscript{28}

One implication of the performance measures of this prediction score is that it could be used as part of scalable efforts to prioritise measures to rehabilitate prisoners, and as part of planned release programmes. This is a consequence of the high negative predictive value at 89%; ie, of individuals identified as low risk, 89% did not in fact reoffend violently within 2 years. Whether this level of accuracy is sufficient for prison services will depend on a range of additional social and political factors, and might, of course, be overridden in individual cases. However, the use of this score provides a framework with which to provide information to make decisions as to which individuals can be released with or without additional service provision. How the 11% failure rate, a corollary of the NPV of 89%, compares with baseline risks in many countries is difficult to establish because violent reoffending risks are not routinely reported. A proxy might be 2-year reinsertion rates that are estimated at 29% in the USA,\textsuperscript{29} and 39% in Australia.\textsuperscript{30}

Different criminal justice and forensic psychiatric systems might take different approaches to such a tool and who should administer it. It could be used by prison health care to help guide community linkage and treatment of prisoners before their release (eg, as part of the Care Programme Approach programme in England or Wales), or by probation services (who typically assess prisoners convicted of severe offences before their release), or case workers who are assigned by some individual US state justice departments to plan sentencing and release arrangements. One strength of our model is that any health-care professional or criminal justice professional can use OxRec. As for timing, this tool could be used towards the end of prison sentences to assist in decisions about the timing of parole and conditions associated with it. Although not validated in community settings, new research could examine its value in guiding probation services to prioritise substance use and mental health interventions in high-risk individuals shortly after prison release.

Several countries are investigating ways to safely reduce prisoner numbers. Notably, in California (USA), a Supreme Court decision in 2014 mandated a reduction in state prisoners,\textsuperscript{11} which will lead to negative consequences if it simply shifts individuals to local prisons. With the probable public health and economic benefits\textsuperscript{31,32} any methods to reduce repeat offending will interest public policy. A second implication of our study is that, because the score’s sensitivity for violent offending at 2 years was 67% at a risk threshold of 20%, it could be used to identify prisoners who could benefit from targeted interventions. If such an intervention is not harmful, such as psychosocial treatments for substance use disorders\textsuperscript{33} or improved links with community-based psychosocial services, then it has the potential to reduce reoffending substantially. Because 88% of the 50% or more risk group had a diagnosis of alcohol or drug use disorders, such treatment could be targeted at these released prisoners. Nevertheless, we acknowledge that such a tool needs to be part of a wider set of risk management strategies in prison, which could include prisoner involvement. These strategies could include more detailed and needs-based assessments. However, the score at this risk threshold should not be used for screening because the specificity (false positive rate) was high at 37%. In other words, three to four of ten offenders were incorrectly identified as being high risk (when in fact they did not reoffend).

Our study has several strengths. Our models were based on a total cohort of released prisoners, with high-quality registers being linked to provide information on covariates and outcomes, including to mortality and emigration registers. Unlike previous studies of risk assessment tools,\textsuperscript{34} we have reported measures of discrimination and calibration, in both derivation and external validation populations. Predictive accuracy was similar in the derivation and the external validation samples; the validation samples were geographically
separated from the derivation samples and done in 10,226 individuals. Alongside the use of prespecified variables and their cutoffs, this method should reduce shrinkage when applied to new populations. Another strength of the method was the use of imputation to replace missing data, which is novel in assessments of risk of violence. Finally, we have provided a web calculator version (OxRec) of the model that is free to use, requires minimal training, and provides a scalable approach to risk prediction.

An important consideration is that the tool has low predictive accuracy at the individual level. It assigns a probability score, similar to risk calculators in cardiovascular medicine such as the Framingham or QRISK score. For the QRISK score, even if a risk threshold is set at 10% for high risk and possible statin use, this means that up to 90% of those will not experience a cardiovascular event in the predicted timeframe. Therefore, one potential harm that is not justified is preventive detention. This prediction score has only been validated in Sweden, and the more different a new prison population is from Swedish prison populations, the more likely accuracy will decrease. For example, the rate of incarceration is relatively low in Sweden but some important prisoner characteristics are similar in Swedish prisoners to other prison populations, such as average length of prison sentence (e.g., 69% of sentences in the Sweden are for <6 months vs 66% in the USA), proportion of prisoners with severe mental illness, and substance use disorder (appendix p 9). Furthermore, US research has shown that substance misuse and mental illness are linked to serious repeat offending. Nevertheless, the tool needs to be validated in countries with different prison populations. The tool also miscalibrated some individuals at the highest risk at 1 year, but this was only based on 18 individuals and was not noted at 2 years. Additionally, our model does not include information on some interview-based risk factors, and institutional misconduct, which might improve the accuracy of the tool, but at the cost of making it more complex and time-consuming.

Our model does not include time-varying covariates intentionally because it aims to provide a snapshot risk score on release from prison, and a different model will probably be required in ex-prisoners, particularly if individuals have been treated for drug and alcohol use disorders. Some specific items such as neighbourhood deprivation might not be generalisable and can be scored as unknown in the web calculator, but the contribution of neighbourhood deprivation to the overall risk score was less than 1%. Diagnoses in this study were restricted to ICD-based clinical ones made by specialists, and some of these might be masked by ongoing substance use. However, the rate of severe mental disorder at 3% in both the derivation and validation samples is similar to the 4% reported in a systematic review of prisoners.

Furthermore, this tool should not be used for forensic psychiatric patients, and might need to be restricted for juveniles, many of whom will stop a cycle of reoffending, and for certain subgroups including minority groups, child sex offenders, and homicide offenders. However, only 1% of the released cohort were homicide offenders, and 2% were sexual offenders; any tool focusing on these outcomes will be limited by low positive predictive values (appendix p 5). Future research could compare the performance of OxRec against currently used instruments in the same sample, and also investigate whether the use of this prediction model reduces risk of violent reoffending in experimental designs.

In summary, the risk score that we have outlined can be used to identify a group of prisoners at high risk of violent reoffending who could be considered for non-harmful interventions, particularly for substance use disorders.

Contributors
SF conceived the study and drafted the manuscript. SF, ZC, NL, TF, and SM designed the methods. PI and HL obtained the data and contributed to study design. ZC did the analyses under the supervision of TF and SM. All authors critically revised the manuscript.

Declaration of interests
NL is the national scientific adviser for Research and Evaluation for the Swedish Prison and Probation Service. All other authors declare no competing interests.

Acknowledgments
This study was funded by grants from the Wellcome Trust (095806), the Swedish Research Council, and the Swedish Research Council for Health, Working Life and Welfare.

References


