Screening for Mental Disorder Using the UK National Offender Assessment System

Probation officers have large caseloads with high levels of psychiatric morbidity but receive minimal training in recognising/managing mental health problems. In the UK, there is no national screening procedure for mental illness among offenders that is considered effective. This study’s aim was to develop a screening system for mental/personality disorders using the Offender Assessment System. Seven screens for mental disorder were developed using items from assessments on 574 prisoners: alcohol misuse, drug dependence, antisocial personality disorder, psychopathy, learning difficulties, psychosis, and severe mood disturbance. Internal validation showed that the screens performed well across accuracy estimates. Prevalence of mental disorders identified with the screens in test and validation samples was similar. However, the psychopathy and severe mood disturbance screens may be more effective in screening for true negatives. The screens are an economical method for prioritising mental health needs of the UK prison and probation population using routinely collected data.

Keywords: mental illness, offender populations, personality disorder, screening

# Introduction

The majority of high risk individuals for future violent and other criminal behaviour are managed by probation officers in the community. Probation officers typically have large caseloads and receive minimal training in recognising and managing mental health problems. The prevalence of mental disorder among offenders receiving community supervision has not been accurately investigated. However, among prisoners released on parole, particularly those receiving longer sentences (Coid et al., 2011), prevalence will be high (Eno Louden, Skeem, & Blevins, 2013; Fraser, Gatherer, & Hayton, 2009; Møller, Stöver, Jürgens, Gatherer, & Nikogosian, 2007; Singleton, Meltzer, Gatward, Coid, & Deasy, 1998; Teplin, Abram, & McClelland, 1996). A meta-analysis showed that prisoners are considerably more likely to have psychosis and major depression, and approximately 10 times more likely to have personality disorder than the general population (Fazel & Danesh, 2002). Furthermore, offenders with multiple co-existing disorders are at higher risk to become violent, report more substance misuse, and more likely to return to custody (Chandler, Peters, Field, & Juliano-Bult, 2004; Ditton, 1999).

Released prisoners with untreated schizophrenia have a threefold risk of violence due to re-emergence of symptoms of their illness (Keers, Ullrich, Destavola, & Coid, 2014). Unfortunately, screening initially on reception to prison (Birmingham, Mason, & Grubin, 1996; Senior et al., 2013)and by healthcare services during sentence (Senior, Appleby, & Shaw, 2014) is largely ineffective and many cases are missed (Shaw et al., 2008). International literature on mental health screens confirm these limitations and have shown that contextual factors such as staff training and ethnicity impact performance (Martin, Colman, Simpson, & McKenzie, 2013). In addition, poor integration between screening used in health and by probation officers have led to negative experiences during screening and a lack of trust by prisoners toward officers (Durcan, 2008). Taken together, more accurate identification of prisoners with poor mental health is necessary to ensure they receive appropriate supervision and intervention together with assessment of associated risks (Chandler et al., 2004).

There is no national screening procedure for mental disorder that is routinely applied among UK offenders and those that are carried out are not always performed consistently and may differ between different locations and offender groups. For example, the health reception screening tool piloted and evaluated by Grubin and colleagues (2002; 1999) was introduced for use throughout prisons in England and Wales. It was developed to detect health problems, particularly serious mental illness. Since its inception, the standardised screen has been modified to meet particular needs of local establishments (Shaw et al., 2008). This contrasts with the application of risk assessments where considerable achievements have been made in their standardisation and routine application.

The National Offender Management Service (NOMS) developed the Offender Assessment System (OASys) to assess risk and needs of adult offenders, assist with management of risk of harm, and link to a supervision or sentence plan. It is routinely carried out by probation officers and used across probation areas and prison establishments in the UK to assess how likely an offender will be reconvicted and the risks that increase that likelihood (Debidin & Fairweather, 2009). Almost seven million prison and probation OASys assessments have been gathered from over one million offenders by the end of March 2014 (Moore, 2015).

The overall aim of the study was to develop a system of screening system for mental and personality disorders using OASys to enable probation officers to identify offenders under supervision in need of treatment and referral for further clinical assessment in addition to their routine assessment of risk. We carried out a longitudinal observational study of a sample of released prisoners who had received OASys and structured clinical assessments to (i) identify OASys items associated with diagnostic categories of mental disorder; (ii) construct screens from these items for seven categories of disorder and test their validity; (iii) test the face validity of the screens in a second independent dataset of released prisoners.

**Methods**

***Participants***

Diagnostic information and corresponding OASys items were obtained from the Prisoner Cohort Study (PCS), which has previously been described in detail (Coid et al., 2007, 2011). In brief, a sample of prisoners in England and Wales (1,396 males and 321 females) were included if they fulfilled the following criteria: (1) serving two years or more for either a violent or sexual principal offence (excluded were life sentence prisoners); (2) at least 18 years of age; and (3) at least six months left to serve (Coid et al., 2007). The sample was interviewed before and after release by twelve research assistants who were trained in all study instruments. Inter-rater reliability was measured interclass-correlations, and these ranged from 0.80 to 0.98 (see Coid et al., 2009). Diagnoses obtained at interview were to act as the “gold standard” against which the diagnostic screens would be tested.

Within the PCS sample, 574 participants (33.43 % of the total N; 543 males and 31 females) had additional OASys ratings carried out independently by probation officers and entered as computer ratings between 15 July 2002 and 7 July 2006. Participants were, on average, 31 years old (SD=11.50, range= 18-75) and primarily Caucasian (n=491, 85.54%). Over half of the sample (n=295) were single before imprisonment.

Additionally, a second independent dataset was utilised to test the face validity of these screens. We were provided with an anonymised sample from NOMS, Ministry of Justice, UK. The sample (N=50,596) was a random selection of general prisoners who were released from prison for the first time between 2005 and 2007. The majority of sample was male (n=43,686, 86.34%), Caucasian (n=39,535, 86.21%), and, on average, 29 years old (SD=9.44, range= 17.03-69.30). The independent dataset included these individuals’ OASys assessments.

***Procedure***

*Outcome measures*

Seven binary outcomes from PCS diagnoses were used to test the accuracy of the screens:

1. Alcohol misuse disorder: participants were screened positive if they scored 16 or above on the Alcohol Use Disorders Identification Test (AUDIT) (Saunders, Aasland, Babor, De La Fuente, & Grant, 1993).
2. Antisocial personality disorder (ASPD): received three or more of 15 items of conduct disorder (CD) prior to age 15 plus three of the seven items defined in DSM-IV indicative of disregard for and violation of the rights of others occurring since age 15 (First, Gibbon, Spitzer, Williams, & Benjamin, 1997).
3. Drug dependence (any): respondents were rated as drug dependent if they were ever dependent on cannabis, stimulants, cocaine, crack cocaine, heroin, opiates, volatile substances, tranquilisers, hallucinogens, ecstasy, or methadone/ physeptone. Dependence was determined by subjective dependence, unable to cut down, tolerance, and withdrawal symptoms.
4. Learning difficulties: the verbal comprehension t-scores of the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999) were used to assess learning difficulties. Two standard deviations below the mean was chosen to represent low IQ.
5. Psychotic disorder: respondents were diagnosed current or lifetime schizophrenia or delusional disorder using the Schedule for Affective Disorders and Schizophrenia-lifetime version (SADS-L) (Endicott & Spitzer, 1978).
6. Psychopathy: the Hare Psychopathy Checklist-Revised (PCL-R) is a 20-item checklist to rate psychopathic traits and antisocial tendencies (Hare, 1991). A cut-off of 27 was applied for categorical diagnosis of psychopathic personality disorder.
7. Severe mood disturbance: participants who scored 11+ on the Hospital Anxiety and Depression Scale for either anxiety and/ or depression (HADS) (Zigmond & Snaith, 1983)*.*

*Development of the screens*

Indicators for the seven categories of mental disorder were identified by an experienced clinician (JWC) who selected OASys items[[1]](#footnote-1) that matched the diagnostic criteria for each category or were considered clinical signs of each disorder (Table 1). Items for psychopathy were matched as close as possible to items in the PCL-R. Each item’s relationship was then tested against whether or not the individual received the diagnosis or outcome of interest. Most OASys items were scored 0 to 2, where “0” indicated no problems and “2” indicated significant problems. A smaller number of chosen items were dichotomous and coded “1” to reflect a negative outcome.

To determine items to be included, we selected those with the highest Area Under the Curve (AUC) values as a simple summary of overall accuracy (Mossman, 2013). The benefit of using AUC is that it is less dependent on base rate of the outcome, and unaffected by changes in sample size (Farrington, Loeber, Stouthamer-Loeber, Van Kammen, & Schmidt, 1996). Our procedure for selecting items for each screen was the following: first, the highest AUC-value from the individual screen items was selected. Second, the next highest AUC-value was selected, and so forth (forward selection), until the overall AUC-value (of the selected items combined) no longer increased.

To ensure we chose the best approach in selecting the items, we examined whether selecting the items in an unadjusted or adjusted logistic regression model or a model based on selecting the highest AUC-value was the best. When testing these different approaches we found no difference in results: the final chosen cut-offs were similar between all three. For the remaining screens, we chose the items based on the latter approach, where we obtained the AUC-values for *all* possible screen-items, and kept the items that contributed to an increase in the overall AUC-value. Table 1 presents the individual items that comprised each screen using this approach.

***Statistical analysis***

In order to validate the screens, we used split sample validation, an approach where the sample is randomly halved. The sample of 574 was randomly split for each screen validation, in which one half of the sample was used for estimation and the other half was used for validation. Once the final items were decided on for each screen (Table 1), they were summed and tested against their respective outcome measure. To develop the screens, cut-off scores were determined through a diverse range of accuracy estimates (McDonald & Calhoun, 2010).

The following estimates were used to decide which cut-off was optimal: (1) sensitivity, the proportion of those who have the diagnosis and who would be screened positive); (2) specificity, the proportion of those who do not have the diagnosis and who would be screened negative; (3) positive predictive value (PPV), the proportion of those who screened positive and who have the diagnosis; (4) negative predictive value (NPV), the proportion of those who screened negative and who do not have the diagnosis; (5) efficiency, also known as the overall hit rate, which is the percentage of those who were correctly identified either as true positives or true negatives; (6) kappa, the strength of agreement between the screen and outcome measure, otherwise known as efficiency adjusted for chance; and (7) estimated population prevalence, the proportion of those in the population who are estimated to have the diagnosis.

In addition to these accuracy estimates, Kraemer and colleagues (1999; 2001) recommended that screens should fulfil two essential requirements: (1) they must significantly be related to the outcome of interest and (2) they must be able to divide a population into high and low risk subgroups, where the prevalence of the outcome is greater in the high risk group. To demonstrate that our screens met these criteria, we included the prevalence of true and false negatives and positives found in a 2x2 contingency table (stratified into those who tested positive and those who tested negative) and the Odds Ratio (OR), which measures the significance and strength of the relationship between the screen and outcome.

The selection of the cut-off point was mainly dependent on the intention of the screen itself. There is no incorrect choice but values should mirror the consequences of incorrect classification made by the instrument (Zhou, Obuchowski, & McClish, 2011). With regards to screening, sensitivity is more important than specificity, as false negatives carry a greater burden than false positives, which may be excluded through further evaluation (Eno Louden et al., 2013). For the final screens, we prioritised a balance between all accuracy estimates.

**Results**

The items selected for inclusion in the screens are listed in Table 1. Overall, the original item pool comprised 87 indicators of mental illness; based on insufficiently high AUC values more than half were discarded. The numbers of items designated to each screen ranged from three to nine.

***Apparent Validation***

Cut-off scores for each screen were chosen based on the preferences described above for the accuracy estimates. Table 2 shows that a cut-off score of three was selected for the majority of the screens, though psychotic disorders had the lowest cut-off score and psychopathy had the highest cut-off score. All AUC-values were above 0.70, and drug dependence achieved the highest value. Additionally, each base rate and estimated population prevalence were relatively similar to each other. When efficiency was adjusted for chance, as measured by kappa, it was lowest for learning difficulties and psychopathy.

About half of the sample was screened positive for alcohol misuse; the screen achieved 0.70 and higher across its accuracy estimates. The base rates for ASPD and drug dependence were also very high (over 50%). The screen for ASPD achieved a similar range of values compared to alcohol misuse, though it was best at identifying 85% of those tested positive as having ASPD. Drug dependence showed over 80% correctly identified for PPV and specificity and about 70% correctly identified for sensitivity and NPV.

Despite learning difficulties, psychotic disorder, and psychopathy correctly identifying less than 55% as positive (sensitivity), they correctly identified a high percentage of negatives as demonstrated by their high NPV and specificity values. Severe mood disturbance had the lowest efficiency, achieving an overall hit rate below 0.70 whereas the highest was achieved by psychotic disorder.

***Internal Validation***

The seven screens were then applied to the other sample to evaluate their performance. As demonstrated in Table 2, with the exception of severe mood disturbance, the AUC-values for each screen were 0.65 or above; the drug dependence screen showed the highest level of discrimination. Sensitivity and PPV values were good for the alcohol misuse, ASPD, and drug dependence screens, showing that these screens were able to identify more than 70% of individuals who had the outcome as positive (sensitivity), and were also able to identify more than 70% of individuals who were tested positive as having the outcome (PPV). The overall percentage correctly classified shows that the first six screens were able to correctly identify over 70% of the sample. Severe mood disturbance performed relatively poorly across all accuracy estimates, with the exception of NPV: it correctly identified 70% of those who did not fulfil the diagnostic criteria.

Table 3 provides further evidence of the screens’ performance. Although the psychopathy screen was significantly related to its outcome, it had a higher number of false negatives compared to true positives. Each screen was significantly related to its outcome for both validation samples with the exception of severe mood disturbance when internally validated. Overall, the screens fulfilled the essential requirements as suggested by Kraemer and colleagues (1999, 2001).

To test the face validity of the screens, an independent and large dataset of released offenders was used. The screens were applied to this dataset and the prevalence of those who were tested positive in each screen was compared with those who were tested positive in the combined validation sample. Table 4 presents the comparison between the PCS combined validation sample and the large general prisoner sample. The prevalence of those who were tested positive for each screen was similar. The overlap of confidence intervals between the PCS and the general prisoner sample further supported their similarities. Only ASPD and psychopathy were significantly more prevalent in the PCS sample (the confidence intervals did not correspond).

**Discussion**

We created seven clinical screens from existing, routinely collected items in the OASys to identify prisoners with the following mental disorders: (1) Alcohol misuse disorder, (2) ASPD, (3) drug dependence, (4) learning difficulties, (5) psychotic disorder, (6) psychopathy, and (7) severe mood disturbance. Application of these screens will help identify prisoners after release in need of further assessment for mental disorder and will help better inform their supervision and intervention following a second stage, skilled assessment of their psychopathology and needs.

The validations demonstrated that the screens have satisfactory validity. When validated, the screens performed moderately well, whereas a few of the screens performed very well in identifying true positives and negatives. The screens showed good discriminatory accuracy, particularly the alcohol misuse and drug dependence screens. Once validated, instruments tend to decrease in their discriminatory accuracy. This was not the case for the alcohol misuse and drug dependence screens.

Each screen was able to correctly identify a higher proportion of those who had the outcome and categorise them in the positively rated group, with the exception of the psychopathy screen. This may either be due to the low base rate of psychopathy or because psychopathy may be difficult to detect with OASys. The psychopathy screen, however, was able to distinguish true negatives, as indicated by its high specificity and NPV: over three-quarters of those who scored lower than 27 on the PCL-R were identified by the screen as negative, whereas almost all of those who were screened negative were true negatives. The screen would therefore be able to provide negative tests with a high degree of confidence. An alternative approach might be to test the psychopathy screen according to its accuracy in identifying ranges of scores on the PCL-R, which we did not investigate in this study. Because the prevalence of psychopathy, however, was lower than the other conditions in the sample, even after increasing prevalence by lowering the cut-off to 27, it was unlikely that greater accuracy would be achieved.

For the severe mood disturbance screen, the validation showed overall poor discriminatory and predictive accuracies. The low accuracy in detecting true positives therefore makes the severe mood disturbance screen less effective as its intended use is to screen those with potential mood disturbance for further evaluation. Detecting low risk individuals may therefore be a better use for this screen because specificity and NPV were moderately high in the validation. For example, as identification of low risk individuals for mood disturbance is high in accuracy, those deemed to be at high risk should receive further assessment.

When we compared the PCS and the general prisoner sample, we found that the screens positively identified similarly between the samples as indicated by the prevalence of those who tested positive. The exception to this was for ASPD and psychopathy, which were found to be higher in the PCS sample than in the general prisoner sample. The suspected reason is that to be part of the PCS sample, individuals needed to fulfil certain criteria (as previously listed), and relative to the general prisoner sample, these requirements characterises this sample as more dangerous. For example, the PCS sample is composed of individuals who were serving two or more years for a violent or sexual principal offence whereas the general prisoner sample is composed of individuals with a variety of offences, ranging from petty to severe offences. In addition, 80% in the PCS sample had either a previous violent or sexual conviction compared to 69% in the general prisoner sample.

***Implications***

The development of the OASys screens for mental and personality disorders was prompted by the need for improved screening and service provision for prisoners, together with the current lack of knowledge regarding psychopathology among their clients on community supervision orders. The OASys, however, was primarily developed for the purpose of assessing risk of future offending. Ratings using the instrument are routinely incorporated into reports for court prior to sentencing, for parole boards, and in sentence planning. For example, Snowden et al. (2007) found that the mandatory Offender Group Reconviction Score (OGRS)(Howard, Francis, Soothill, & Humphreys, 2009), an instrument within the OASys that assesses the likelihood of recidivism, did not accurately predict recidivism in mentally disordered offenders, highlighting the need for assessments that accounted for mental illness. Similarly, Hickey, Yang, and Coid (2009) found that diagnosis of mental disorder was an important component of risk assessment using the PCS sample, and that ability to accurately predict future violent reoffending was seriously compromised by ASPD and psychopathy.

Future research may show important effects of positive screening on probation officers’ accuracy of assessing risk using the OASys with implications for subsequent risk management. Because the mental health of prisoners has a direct bearing on these aspects of assessment, prison officers and prison staff need to know whether to seek advice from mental health professionals. As the majority of cases of severe mental disorder are not detected whilst in prison, and because there is no information on the prevalence of mental disorder among those serving community sentences, these screens have service implications, as they provide an opportunity for further evaluation by mental health professionals on whether the offender has the condition for which they have screened positive. OASys ratings are carried out routinely, beyond reception, and even after release. These tests are therefore inexpensive, easy to administer, and impose minimal discomfort. As they are computerised, providing results on positive or negative screening could be automated.

***Limitations***

The risk and benefits of the new screening procedure we have developed needs to be externally validated before introduction into routine use within the criminal justice system, together with policies for information sharing with health care professionals. In addition, the split sample method used to validate the screens have its own limitations, as the use of only half of the sample may result in biased estimates (Steyerberg et al., 2001). Further validations of these screens are needed.

Screening is more appropriate for serious clinical conditions, the prevalence should be high among the population screened, and intervention should ideally be before the condition develops. Whereas the first two factors clearly apply, the screens have been standardised on conditions that are already well established. Furthermore, there are potential difficulties in screening positive for conditions such as ASPD and psychopathy which are associated with stigma, therapeutic pessimism, and where in the prison system it has been debated as to whether psychopaths should receive treatment, whether treatment improves their condition, or whether it makes the risks of criminal recidivism worse (Hare, 1993; Seto & Barbaree, 1999; Skeem, Monahan, & Mulvey, 2002). Nevertheless, there is increasing research into the interactions between mental disorder and risk factors for future offending.

The clinical screens developed in this paper were specific to UK prison populations. The OASys applies only to the UK probation and prison services, and therefore, the application of these screens is limited to the British context. However, improving the mental health of prisoners is a universal issue.

***Conclusions***

The implicit assumption underlying screening is that early detection before development of symptoms (or further offending behaviour in the case of OASys ratings) will lead to a more favourable prognosis because treatment started before the disease became clinically manifest and is therefore more effective than later treatment. In the case of schizophrenia, the need to identify prisoners before release who have not received treatment because they have not been identified by in-reach prison services is now apparent (Keers et al., 2014). However, because the OASys contains few if any items that are symptoms of severe mental illness, the screens are heavily dependent on individuals who have been identified as needing treatment with the risk that cases at highest risk of future violence may be missed. Improvement in this area would require the addition of further items to identify symptoms in an already extensive risk assessment. Nevertheless, self-report questions administered by lay interviewers without clinical training have proved successful in identifying cases with psychotic illness both in prisons and the community (Bebbington & Nayani, 1995; Singleton et al., 1998).

Further research is needed in two main areas. First, studies to evaluate screening in offender populations using the measures we have developed in an attempt to demonstrate a relationship between screening and decline in morbidity or mortality due to increased use of mental health services. Second, whether use of the screens can improve accuracy of future estimates of risk of reoffending when combined with routine assessments of risk, and more importantly guide risk management and interventions designed to mitigate risk.

References:

Bebbington, P., & Nayani, T. (1995). The psychosis screening questionnaire. *International Journal of Methods in Psychiatric Research*, *5*(1), 11–19.

Birmingham, L., Mason, D., & Grubin, D. (1996). Prevalence of mental disorder in remand prisoners : Consecutive case study. *British Medical Journal*, *313*(7071), 1521–1524.

Chandler, R. K., Peters, R. H., Field, G., & Juliano-Bult, D. (2004). Challenges in implementing evidence-based treatment practices for co-occurring disorders in the criminal justice system. *Behavioral Sciences and the Law*, *22*(4), 431–48.

Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Roberts, A., Roberts, C., … Farrington, D. P. (2007). *Predicting and understanding risk of re-offending: The Prisoner Cohort Study*. London: Ministry of Justice.

Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Farrington, D. P., & Rogers, R. D. (2011). Most items in structured risk assessment instruments do not predict violence. *Journal of Forensic Psychiatry & Psychology*, *22*(1), 3–21.

Coid, J. W., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., … Rogers, R. D. (2009). Gender differences in structured risk assessment: comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, *77*(2), 337–48.

Debidin, M., & Fairweather, L. (2009). Introduction to the OASys and research on OASys 2006 to 2009. In M. Debidin (Ed.), *A compendium of research and analysis on the Offender Assessment System (OASys) 2006-2009* (pp. 1–9). London: Ministry of Justice.

Ditton, P. M. (1999). *Mental Health and Treatment of Inmates and Probationers*. Washington DC: U.S. Department of Justice, Office of Justice Programs.

Durcan, G. (2008). *From the inside: Experiences of prison mental healthcare.* London: Sainsbury Centre for Mental Health.

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, *35*(7), 837.

Eno Louden, J., Skeem, J. L., & Blevins, A. (2013). Comparing the predictive utility of two screening tools for mental disorder among probationers. *Psychological Assessment*, *25*(2), 405–15.

Farrington, D. P., Loeber, R., Stouthamer-Loeber, M., Van Kammen, W., & Schmidt, L. (1996). Self-reported delinquency and a combined delinquency seriousness scale based on boys, mothers, and teachers: Concurrent and predictive validity for African-Americans and Caucasians. *Criminology*, *34*(4), 493–517.

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23000 prisoners: A systematic review of 62 surveys. *Lancet*, *359*(9306), 545–50.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B., & Benjamin, L. (1997). *Structured clinical interview for DSM-IV Axis II Personality Disorders, (SCID-II)*. Washington DC: American Psychiatric Press, Inc.

Fraser, A., Gatherer, A., & Hayton, P. (2009). Mental health in prisons: Great difficulties but are there opportunities? *Public Health*, *123*(6), 410–4.

Grubin, D., Carson, D., & Parsons, S. (2002). *Report on New Prison Reception Health Screening Arrangements: the results of a pilot study in ten prisons*. Newcastle Upon Tyne: University of Newcastle.

Grubin, D., Parsons, S., & Hopkins, C. (1999). *Report on the evaluation of a new reception health questionnaire and associated training*. London: HM Prison Service.

Hare, R. (1993). *Without conscience: The disturbing world of the psychopaths among us*. New York: Guilford Press.

Hare, R. D. (1991). *The psychopathy checklist—revised (PCL–R)*. Toronto: Multi-Health Systems.

Hickey, N., Yang, M., & Coid, J. W. (2009). The development of the Medium Security Recidivism Assessment Guide (MSRAG): An actuarial risk prediction instrument. *Journal of Forensic Psychiatry & Psychology*, *20*(2), 202–224.

Howard, P. D., Francis, B., Soothill, K., & Humphreys, L. (2009). *OGRS 3: The revised offender group reconviction scale*. London: Ministry of Justice.

Keers, R., Ullrich, S., Destavola, B. L., & Coid, J. W. (2014). Association of violence with emergence of persecutory delusions in untreated schizophrenia. *The American Journal of Psychiatry*, *171*(3), 332–9.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1999). Measuring the potency of risk factors for clinical or policy significance. *Psychological Methods*, *4*(3), 257–271.

Kraemer, H. C., Stichare, E., Kazdin, A. E., Offord, D. R., & Kupfer, D. J. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry*, *158*, 848–856.

Martin, M. S., Colman, I., Simpson, A. I. F., & McKenzie, K. (2013). Mental health screening tools in correctional institutions: a systematic review. *BMC Psychiatry*, *13*, 275.

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: A critical review. *Clinical Psychology Review*, *30*(8), 976–87.

Møller, L., Stöver, H., Jürgens, R., Gatherer, A., & Nikogosian, H. (2007). *Health in prisons: A WHO guide to the essentials in prison health*. Copenhagen: World Health Organization.

Moore, R. (2015). *A compendium of research and analysis on the Offender Assessment System (OASys) 2009-2013*. London: Ministry of Justice.

Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences and the Law*, *39*(January), 23–39.

National Offender Management Service. (2010). *O-DEAT data post-OASys Release 4.3.1 (end of August 2009): Explanatory notes*. London: Ministry of Justice.

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804.

Senior, J., Appleby, L., & Shaw, J. (2014). The management of mental health problems among prisoners in England and Wales. *International Psychiatry*, *11*(3), 56–58.

Senior, J., Birmingham, L., Harty, M. A., Hassan, L., Hayes, A. J., Kendall, K., … Shaw, J. (2013). Identification and management of prisoners with severe psychiatric illness by specialist mental health services. *Psychological Medicine*, *43*(7), 1511–20.

Seto, M. C., & Barbaree, H. E. (1999). Psychopathy, treatment behavior, and sex offender recidivism. *Journal of Interpersonal Violence*, *14*(12), 1235–1248.

Shaw, J., Senior, J., Hassan, L., King, D., Mwasambilli, N., Lennox, C., … Weston, J. (2008). *An evaluation of the reception screening process used within prisons in England and Wales*. Manchester: Offender Health Research Network.

Singleton, N., Meltzer, H., Gatward, R., Coid, J. W., & Deasy, D. (1998). *Survey of Psychiatric Morbidity Among Prisoners in England and Wales*. London: Office for National Statistics.

Skeem, J. L., Monahan, J., & Mulvey, E. P. (2002). Psychopathy , treatment involvement , and subsequent violence among civil psychiatric patients. *Law and Human Behavior*, *26*(6), 577–603.

Snowden, R. J., Gray, N. S., Taylor, J., & MacCulloch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine*, *37*(11), 1539–49.

Steyerberg, E. W., Harrell, F. E., Borsboom, G. J. J. ., Eijkemans, M. J. ., Vergouwe, Y., & Habbema, J. D. F. (2001). Internal validation of predictive models. *Journal of Clinical Epidemiology*, *54*(8), 774–781.

Teplin, L. A., Abram, K. M., & McClelland, G. M. (1996). Prevalence of psychiatric disorders among incarcerated women. *Archives of General Psychiatry*, *53*(6), 505–12.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI) manual*. San Antonio: The Psychological Corporation.

Zhou, X.-H., Obuchowski, N. A., & McClish, D. K. (2011). *Statistical methods in diagnostic medicine* (2nd ed.). New Jersey: John Wiley & Sons.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–70.

1. Several OASys items used in the study have been discontinued as of 2009 (National Offender Management Service, 2010) : violent behavior related to drug use (drug dependence); convictions for burglary (adult criteria for ASPD); breach probation/ parole and five or more convictions (psychopathy). [↑](#footnote-ref-1)