Evaluation of an Enhanced Behaviour Monitoring system in UK open prisons

Gary Goodley[ab]* and Dominic Pearson (ORCID ID 0000-0002-4766-6841)[b]

[a]*HMPPS Psychology Services, HMP North Sea Camp, Boston, UK;*
[b]*Department of Psychology, University of Portsmouth, Portsmouth, UK*

*Correspondence to G. Goodley at gary.goodley@justice.gov.uk, or at HMP North Sea Camp, Croppers Lane, Freiston, Boston, PE22 0QX

Accepted for publication in *Criminology & Criminal Justice* on 6th October 2023

Gary Goodley, PhD is a practitioner forensic psychologist working in Her Majesty's Prison and Probation Service. His research is focused on factors associated with recidivism, risk management in open prisons, and behavioural monitoring as a tool for risk identification and management.

Dominic Pearson, PhD is an Associate Professor in Clinical Research and Practice in Forensic Psychology at the University of Portsmouth, UK. He is a registered practitioner forensic psychologist with over 20-years' experience with the prison and probation services, studying community-oriented approaches to correctional practice; including the constructive use of behaviour monitoring.

Word count: 8828

Evaluation of an Enhanced Behavioural Monitoring system in UK open prisons

Abstract

Behavioural monitoring has efficacy in predicting recidivism. As an intervention however, the proven effectiveness is limited. The present study is an evaluation of the Enhanced Behaviour Monitoring (EBM) scheme operating across open prisons in England and Wales, to reduce instances of failure (e.g., abscond, reoffending). Utilising a matched sample design, logistic regression analyses showed EBM had null effects on serious recidivistic outcomes (e.g., abscond, re-offending). Those allocated to EBM were more likely to get recalled before completing the intervention. In essence, EBM was utilised primarily as a surveillance programme – to defend against risk exposure – resulting in recalls to closed prisons misaligned to the community recidivism risk; undermining the open prison concept. We discuss the findings in the context of the adverse effects of 'surveillance only' community supervision programmes. We argue for developing the rehabilitative potential within EBM and mitigating against *excessive* risk exposure via systems-level policy change.

Keywords: Abscond; Behavioural Monitoring; Open prisons; Recidivism; Risk management; Temporary Release Failure

The presumptions underlying the most commonly used forensic risk assessment tools are majority-biased and result in the misclassification of a significant proportion of offenders (Eckhouse et al., 2019; Monahan et al., 2017). Eckhouse and colleagues (2019) argued that data-driven risk prediction tools are rarely individualised to the offender and assessments utilising group-based averages discriminate against minority groups. One group where the risk of recidivism is often overestimated is older offenders (Monahan et al., 2017), such that the accurate assessment of those serving lengthy indeterminate sentences for serious offences has been considered a long-standing problem for criminal justice systems (cf. Clark et al., 1993). Conversely, risk underestimation is common in some groups, such as a subset of offenders with limited criminal histories who later go on to commit homicide (Greenall & Richardson, 2015; Soothill et al., 2002). It is these cases which, particularly where a further offence is the outcome, have the most damaging consequences, causing public outrage, and increased scrutiny of those organisations managing the risk (cf. Dawar & Davis, 2014).

Some academics argue that forensic risk assessment – a *prognostic* task (i.e., prediction of a future event) – should not be held to the evaluative standards of *diagnostic* assessments (i.e., identification of a condition that already exists) (Helmus & Babchisin. 2017). For instance, detecting whether an individual has cancer (yes or no) is subject to much less uncertainty than detecting whether an individual will get cancer, which can only be expressed in probabilistic terms (i.e., percentage chance based on a

multitude of known and unknown risk-increasing factors). In contrast, rather than relying on group-based probabilities, others propose behavioural monitoring as a promising means of individualising risk assessment, contemporaneously forecasting reoffending probability, and optimising risk management (Clark et al., 1993; Goodley et al., 2022; Jones, 2004; Pearson & McDougall, 2017). In a behavioural scheme, it is possible to identify behaviour patterns at the time of offending and make risk judgements based on the present 'flow of behaviour' (Jones, 2000). The review below suggests behavioural monitoring protocols have potential to structure professional judgements of risk and monitor progress at a more nuanced level, thereby facilitating intervention following the observation of risk-related behaviour. This article presents an evaluation of a behavioural monitoring protocol – Enhanced Behaviour Monitoring (EBM) (NOMS, 2015) – implemented nationally across open prisons in England and Wales since 2015.

The Utility of Behaviour Monitoring

Behavioural monitoring has the potential to overcome some of the systematic constraints observed in forensic risk assessment. For instance, there is growing recognition in the literature that the invariant application of standardised risk factors within both actuarial and Structured Professional Judgement (SPJ) tools, may result in bias against certain cultural groups (Shepherd & Lewis-Fernandez, 2016; Schmidt et al., 2019; Venner et al., 2021). Indeed, the majority of these tools are normed on Western

populations – and their associated world views and norms – creating disparities in the accurate assessment of risk in out-of-sample minority groups (Shepherd & Lewis-Fernandez, 2016). Conversely, behavioural monitoring protocols can be tailored to the individual client and behavioural aberrations can be assessed in a way which is culturally sensitive (Schmidt et al., 2019). Thus client-worker discussions arising from behaviour monitoring can help make explicit any cultural issues in supervision, and so contribute to culturally sensitive practice.

Likewise, some academics argue that risk factors within SPJ tools are susceptible to misclassification as a function of time/resource restrictions (Ho et al., 2018; Vojt et al., 2013). By their very nature, dynamic risk assessments have short-term validity, necessitating regular review. Yet the administration of SPJ tools is time-consuming (Green et al., 2010); the accurate discrimination of case-relevant factors is impeded by time-constraints resulting, arguably, in flawed decision-making (Ho et al., 2018). In the absence of regular review, limitations also emerge in the way risk factors are measured and evaluated. Jones (2004) argues that conventional risk assessment protocols often result in risk predictions based on past, discrete episodes of high-level behaviour. Harris and Rice (2015) assert that dynamic risk factors become static in nature when measured once: the dynamic risk information – the very aspect that makes the risk factor changeable (Douglas & Skeem, 2005) – becomes lost. In contrast, behavioural consistency can be measured more accurately and effectively, over both time and

context, using a behaviour monitoring scheme (Clark et al., 1993; McDougall et al., 2013; Pearson & McDougall, 2017).

Behavioural monitoring systems necessitate staff training in the identification of risk-related behaviour and are premised on effective inter-agency communication. As such, there is scope to delve deeper than official, often incomplete, information sources (cf. McDougall et al., 2013; Pearson & McDougall, 2017). Institutional misconduct offences – as an official risk information source – provides a case in point. 'Situational muting' (Daffern et al., 2007) – the suppression, by the prison environment, of the prototypical precipitants of the offending behaviour and the limiting of access to victims/opportunities to offend, might diminish expressions of antisocial behaviour (Gordon & Wong, 2010; Jones, 2004). Likewise, some institutional infractions simply go unrecorded in wing behaviour records rendering official information sources, as an indicator for violence propensity, unreliable (Adams, 1992; Mooney & Daffern, 2015). The official recording of incidents is dependent on the recording threshold of individual members of staff (Bottoms, 1999). Prison officers are the gatekeepers of observed behaviour; but through a combination of normalisation (e.g., desensitisation to violence), procedural (e.g., insufficient time, lack of observable consequences) and individual (e.g., concerns about defensibility, failure to recognise the significance of the behaviour) factors, important risk-related information goes unreported (Atkinson & Mann, 2012). By engaging offenders and inter-agency staff in the risk assessment

process and tapping into observations of ‘daily prison life’, behaviour monitoring can unlock current risk-related information, inaccessible to other risk assessment methodologies.

Utilising Behaviour Monitoring Effectively

Clark and colleagues (1993) were pioneers in creating a simple behaviour monitoring framework applicable to prisons, later known as the Wakefield Behavioural Risk Assessment model (McDougall et al., 1995). Using behavioural analysis techniques, they were able to identify offence-related behaviour similar to the index offence in 60% of prison behaviours. Building on these findings, McDougall and colleagues (2013) developed the ADViSOR methodology to examine the reverse: whether the community behaviour of those offenders at high risk of committing a violent and/or serious sexual offence on release from prison, could be predicted from their custodial behaviour. However, rather than solely assessing consistency between the index offence and prison behaviours as per Clark et al. (1993), they looked at prison behaviour more holistically, recording positive and negative behaviours not necessarily linked to the index offence nor functionally equivalent. The study also went beyond those behaviours recorded in official records, recruiting specifically trained staff to record any behaviour of concern observed across custodial environments (e.g., workshops, education). They found that the frequency of concerning behaviours in custody and the community were significantly correlated and that return to custody for

re-offence or recall, could be predicted by the frequency of such behaviours observed in custody. Not only was the prediction of recidivism highly accurate (92%), but there was also qualitative similarity in behaviours observed across prison and the community with 80% of behaviours rated as 'similar' or 'very similar'. McDougall et al. (2013) argue that the prison environment does not suppress risky behaviour entirely but rather the expression may change; ensuring staff are trained to spot these subtle behavioural nuances is essential to effective risk monitoring.

Behaviour Monitoring as an Intervention

The Clark et al. (1993) and McDougall et al. (2013) studies were primarily interested in identifying whether offence-related behaviour is cross-situational but stopped short of considering whether behaviour monitoring, via the collection of personalised risk information, could be used to intervene; to divert individuals from offending behaviours. There are, however, studies in existence to suggest that this is worth pursuing. Simpson and colleagues (2015) demonstrated potential to reduce the number of absconding incidents in a secure forensic psychiatric sample purely through policy change, i.e., the adoption of a risk assessment protocol guided by structured professional judgement principles and multidisciplinary decision-making. The abscond rate reduced from 17.8% pre-implementation to 12.0% post-implementation when release decisions considered risk indicators as outlined in the Historical-Clinical-Risk-20 (HCR-20); when the temporary release was purposeful (i.e., facilitated

rehabilitation/reintegration); and when exposure to risk was gradual, with leave increasing in frequency and intensity over time. Tolisano and colleagues (2017), also using a forensic psychiatric sample, demonstrated that it was possible to modify violent behaviour using risk assessment as the starting point for a wrap-around support package. Presenting two case studies, they showed a reduction in instances of aggression and self-injury over time when a behavioural support plan was informed by a functional behavioural analysis (FBA). The FBA enabled the staff team to identify and remove mismatches between the individual's needs and the environment, coach positive alternative behaviours, and use effective behavioural strategies such as positive reinforcement to influence behaviour change. Similarly, using a risk assessment protocol as the starting point to determine offence-related risk factors, associated skills deficits and environmental triggers, Ward and Bosek (2002) were able to tailor a community-based behavioural risk management plan to a group of adolescent and adult males with developmental disabilities engaging in inappropriate sexual behaviours. Once more, the risk management plan consisted of ensuring application of robust monitoring and supervision systems whilst coaching positive alternative behaviours. None of the 41 individuals in their sample re-offended during the monitoring period.

Whilst the evidence for using behavioural monitoring as the basis for intervention is encouraging, caution is heeded from the intensive supervision literature. Intensive supervision – a form of community supervision employing more frequent contacts and a

variety of other mechanisms to increase surveillance and control (Barnes & Hyatt, 2018), impacts positively on recidivism outcomes only when supervisory approaches incorporate a therapeutic element or supervisors act as 'agents of change' (Bonta et al., 2011; Drake, 2018; Petersilia & Turner, 1990; Taxman, 2008). In the absence of these elements, such approaches are ineffective at reducing crime and may have iatrogenic effects, serving only to detect non-compliant offenders and increase the rate of reincarceration due to technical violation (Drake, 2018; Hyatt & Barnes, 2017; Petersilia & Turner, 1993). Indeed, Paparozzi and Gendreau (2005) found that rehabilitation-orientated parole officers who revoked licences sparingly, yielded greater reductions in recidivism including new convictions, compared with law enforcement-orientated parole officers. However, Clear and Latessa (1993) suggest that the pattern may be more nuanced. They found that officer role preference is mediated by the philosophy of the organisation such that officers, regardless of role preference, were more likely to select supportive tools to manage the risk when in organisations with a rehabilitative philosophy. Goodley and Pearson (2023) have previously shown that prison managers in open prisons, given the serious implications of adverse events such as abscond and re-offending, take a conservative approach to non-compliance, misaligned to the community re-offending risk. Intervention supported by behavioural monitoring presents an opportunity, where implemented appropriately, to redress this balance.

The Current Study

The current study aimed to evaluate a behaviour monitoring protocol – Enhanced Behaviour Monitoring (EBM). EBM is a policy framework implemented in 2014 in open prisons across England and Wales to mitigate instances of abscond, temporary release failure, and serious further offending by resident inmates (NOMS 2015). The process is detailed in NOMS (2015) but in brief, EBM consists of a psychologist-led file review, and for those deemed at heightened risk for failing, a formal process enabling those involved in the offender's management to observe and share observations related to the individual's behaviour. The psychologist-led file review - recognising the importance of a detailed and expert analysis of behaviours indicative of risk manifestation, as per previous studies (Clark et al., 1993; McDougall et al., 2013) – provides the basis for intervention. Global behaviours indicative of risk manifestation are listed in a proforma and provide the yardstick against which the individual is assessed. These assessments are led primarily by the prisoner's offender manager – based on the behavioural observations of staff involved in the prisoner's management – and take place over a period of six once-monthly meetings with the prisoner. The meetings provide a behavioural feedback loop to the prisoner and an opportunity set goals and coach the prisoner to mitigate emerging risks. Upon completion of the six-month period, having gained assurance that behaviour is stabilised, the standard

expectation is that the prisoner returns to ordinary prison management. The present evaluation, using a matched sample, set out to examine whether EBM is effective at reducing 'failure' rates. It was hypothesised that those subject to the behaviour monitoring intervention would be less likely to 'fail' despite the additional monitoring.

Method

Design and Procedure

A quasi-experimental design was used to evaluate the effectiveness of Enhanced Behavioural Monitoring (EBM) on failure outcomes in a sample of adult male prisoners resident within open prisons within England and Wales. Those eligible for residency in open prisons in England and Wales are those who have less than three years left to serve, or less than three years until their parole release eligibility date. They must have been assessed as a low risk of abscond, a low risk of harm to the public, and as unlikely to continue to engage in criminality (Ministry of Justice, 2020). Open prisons in England and Wales typically have no perimeter security, and apply, compared to a closed prisons, fewer physical (e.g., gates, locks, fences) and procedural (e.g., rules, regulations) controls. Prisoners are typically afforded graduated opportunities to engage in resettlement activities in the community such as finding or accessing education, training, or work; rebuilding family ties; and familiarisation with their resettlement area (Ministry of Justice, 2019).

The dataset consisted of 692 prisoners all of whom were eligible for EBM, meeting one or more of the following criteria as per Prison Service Instruction 13/2015 (NOMS, 2015): serving an indeterminate sentence; subject to Multi Agency Public Protection Arrangements (MAPPA); and/or, assessed as high or very high risk of harm on the Offender Assessment System (OASys) risk assessment tool (Home Office, 2006). All were subject to a case file review between 11 June 2014 and 21 November 2017 with those deemed at heightened risk for failure being progressed for behaviour monitoring intervention, as described above.

To ensure sufficient numbers of men in the intervention group, data were collected using a purposive sampling method. Allocation to the intervention group was determined by recommendation for management under EBM – akin to an intention to treat (ITT) approach. Of the total dataset (N = 692), 174 prisoners had a formal recommendation for management under EBM. The remaining 518 prisoners were managed under existing protocols and represented the pool of prisoners from which the control group was selected, using the matching procedure described below.

Demographic and offence-related data were collected from data sources including the EBM case file review, EBM database, National Offender Management Information System (NOMIS), and public protection unit database. Data included age, current offence, sentence type, number of previous convictions, personality disorder diagnosis, previous failures in open conditions, past prison behaviour (behaviour in the one-year

prior to transferring to open conditions; total adjudications), recent prison behaviour (substance misuse, adjudications, and adverse events in the last six months), and a current risk assessment score (e.g., OASys Violence Predictor).

Measures

The outcome against which the EBM and control group were compared was the observation of a 'failure' event. A failure was recorded when any of the following were observed: abscond, custodial re-offence, temporary release failure (TRF), or security recall. Abscond was defined as unlawfully gaining liberty without overcoming a physical security restraint. Custodial re-offence was defined as reconviction for any offence, whilst resident in open conditions. TRF was defined as a failure to adhere to any condition included on the individual's temporary release licence. Security recalls were recalls to secure conditions, initiated by a prison manager, typically following a breach of the prison rules, deemed indicative of an intolerable increase in the individual's risk. Data were retrieved from the NOMIS database and cross-referenced against other records to validate the failure. In a small number of cases, the custodial failure could not be validated - the decision to fail was clearly overturned within a 28-day period and the individual returned to open conditions. Such cases were not classified as a failure and the follow-up period continued until the point of release or recording of a valid failure. Data pertaining to custodial 'failure' and community

recidivism were collated in March 2022. All participants had either graduated or 'failed' in open conditions at the end of the follow-up period.

Matching Strategy

To ensure the control group matched the EBM group as closely as possible, we employed Propensity Score Matching (PSM) (Rosenbaum & Rubin, 1983). PSM is a quasi-experimental matching technique designed to overcome selection bias when estimating the effect of a given intervention. The possibility of bias occurs because a difference in the outcome between the intervention and control groups may be caused by a factor which predicts assignment to the intervention group rather than the impact of the intervention itself. PSM aims to minimise this bias by accounting for covariates that predict assignment to the intervention group – in this case, failure in open conditions. This is achieved through a process of 'matching' individual subjects in the intervention and control groups according to these covariates, achieving 'balance' between the two groups (Austin, 2011; Rosenbaum & Rubin, 1983). Through this process of matching, pre-existing differences between the groups can be minimised meaning any observed differences in outcomes can, with greater certainty, be attributed to the effect of the intervention (Stuart, 2010).

To ensure the EBM and control groups were matched on those covariates that may confound the relationship between the intervention and outcome, we broadly followed the covariate selection method outlined in Lee and Little (2017). First, and based on the

findings of Goodley and Pearson (2023), we identified those covariates related to the outcome variable – in this case, failure in open conditions. Goodley and Pearson (2023) identified five predictors of failure in open prisons in England and Wales, namely: adjudications in the six months prior to outcome, substance misuse in the six months prior to outcome, behavioural warnings in the one-year prior to transfer to open conditions, personality disorder diagnosis, and OASys Violence Predictor (OVP) score. From this pool of covariates, we discarded adjudications and substance misuse in the six months prior to outcome given they were susceptible to influence by the intervention itself. The three remaining variables were retained as the matching criteria.

The EBM and control groups were matched using nearest neighbour matching. In nearest neighbour matching, an individual in the intervention group is matched to the individual in the control group with the most similar or nearest propensity score (Rosenbaum & Rubin, 1985). Consequently, this method can result in individuals with relatively dissimilar propensity scores being matched. To avoid 'bad matches', we specified a caliper distance – a maximal acceptable distance in which propensity scores can be matched. Austin (2011) suggests using a caliper width equal to .2 of the standard deviation of the logit of the propensity score to eliminate approximately 98-99% of the bias caused by measured confounding variables. A corresponding caliper of .04 was imposed on the data. We also employed nearest neighbour matching with replacement, meaning that individuals could be matched multiple times if their

propensity score was within the least distance to more than one individual in the corresponding group (Dehejia & Wahba, 2002). Three members of the EBM group could not be matched to a member of the control group and were removed from the analyses. In total, there were 171 offenders in the intervention group and 171 in the control group. Table 1 shows the sample characteristics pre- and post- matching, including the mean scores for the variables upon which the intervention and control groups were matched. Logistic regression analyses confirmed the groups were adequately matched as group membership could not be differentiated based on these variables: Behavioural warnings χ^2 (1, N = 342) = 1.30, p =.256; OVP score χ^2 (1, N = 342) = 0.13, p =.717; Personality disorder diagnosis χ^2 (1, N = 342) = 3.46, p =.327. Overall, compared to the full sample, the EBM and matched control groups had a higher number of behavioural warnings prior to transferring to open conditions, higher risk scores, and greater prevalence of personality disorder – indicative of a higher level of problematic behaviour in these groups – placing them at greater risk of failure, as per Goodley and Pearson (2023).

Table 1. *Sample Characteristics*

Variable	**Full sample *(N = 692)***	**EBM group (*N = 171)***	**Control group (unmatched) (*N = 518*)**	**Control group (ITT-matched) (*N = 171*)**
Age, M_{years} *(SD, Range)*	39.0 *(12.6, 19-81)*	37.5 *(11.2)*	39.6 *(13.0)*	35.6 *(10.7)*
Previous convictions, M_{number} (*SD, Range*)	9.7 *(10.9, 0-89)*	15.0 *(14.8)*	7.8 *(8.6)*	13.2 *(10.7)*
Index offense (%)				
Burglary	5.9	8.3	5.1	9.8
Drug	2.5	1.2	2.9	2.4
Fraud	1.0	0.0	1.4	0.0
Robbery	16.6	23.7	14.3	19.5
Sexual	29.4	21.9	31.6	20.1
Violence	38.3	37.3	38.7	41.5
Other	6.3	7.6	6.0	6.7
Sentence type (%)				
Determinate	42.5	21.6	49.4	44.4
Indeterminate for Public Protection	31.6	46.8	26.4	35.7
Life	25.9	31.6	24.1	19.9
Behavioural warnings in 12 months prior to transfer to open conditions M_{number} (*SD, Range*)	2.1 *(3.1, 0-19)*	4.1 *(3.98)*	1.5 *(2.3)*	3.6 *(3.6)*
Personality Disorder Diagnosis				
Not assessed	59.2	48.5	62.9	39.8
Negative diagnosis	13.0	12.3	13.3	17.0
Subthreshold traits	19.1	26.3	16.4	31.0
Positive diagnosis	8.7	12.9	7.3	12.3
OASys Violence Predictor (OVP) score M_{number} (*SD, Range*)	26.7 *(15.9, 2-85)*	34.6 *(15.5)*	23.9 *(15.0)*	35.2 *(17.4)*

Analytical Strategy

Using an ITT approach, the aim was to determine the effect of EBM on failure rates in the open prison estate. A sequential logistic regression analysis was performed to determine the predictive ability of group membership (EBM vs. control). Given residency in open prison is time-limited as a function of sentence-type/length, which varied across the sample, it was necessary to estimate and control for the time 'at risk' in open conditions (release date minus arrival in open conditions date). This was necessary given the potential bias for instance, of comparing the likelihood of survival for prisoners serving longer compared to shorter sentences given the additional time 'at risk' of failure in open conditions. For indeterminate prisoners who failed in open conditions, this date was estimated based on their next scheduled parole hearing. Given actual time at risk is subject to the impact of failure/recall events, it was necessary to generate an 'expected' time at risk (i.e., expected release date minus arrival in open conditions date). Expected time at risk ranged from 48 to 2,214 days with a Mean expected stay of 495.67 days (*SD* = 263.53). Actual time at risk ranged from 20 to 2,214 days with a Mean actual stay of 364.72 days (*SD* = 299.46). At Block 1 therefore, the expected time-at-risk variable was entered into the analysis, followed by the group variable at Block 2. We also performed Area Under the receiver-operating characteristic Curve (AUC) analyses on the data to test the performance of the model,

and Kaplan-Meier tests to compare the survival rates of the EBM and control group to determine whether EBM had any effect on longevity in open conditions.

ITT is an effective method for drawing conclusions about the attempt to intervene – exposing problems in referral procedures and programme retention but, arguably, has limitations in assessing the effects of the intervention actually received (cf., Hollin, 2008; Kovach, 2020). The intervention group is not only inclusive of those who were discharged from EBM management per-protocol (i.e., completers) but also those who failed before starting (i.e., non-starters) or completing (i.e., non-completers). These groups potentially dilute any positive effect of EBM. As such, we also adopted a treatment received (TR) approach, repeating the above sequential logistic regression analysis, to understand whether completion of EBM per-protocol had any discernible impact on failure rates compared with a matched sub-group of control cases.

Results

Intention to Treat (ITT) Analysis

Univariate Analysis. The majority (n = 211 [61.7%]) of the sample failed in open conditions. Except for four failures which could not be verified, 23 of 211 absconded (10.9%), five re-offended (2.4%), and five breached the terms of their temporary release licence (2.4%). The remaining and most significant proportion of the failures were security recalls (n =174; 82.5%). Recall was initiated after a single adjudication on

average (M = 0.69, SD = 1.07). Of the sample released into the community for at least one year (n = 272), 20 (7.4%) reoffended and 58 (21.3%) were recalled to custody.

The EBM group failed in open conditions more frequently (n = 126) than the ITT-matched control group (n = 85). The difference was statistically significant χ^2 (1, N = 342) = 20.80, $p < .001$. However, the number of serious recidivistic events did not differ across the samples with 12 (7.1%) of the EBM group absconding (ITT control group: n = 11; 6.5%), three (1.8%) re-offending (ITT control group: n = 2; 1.2%) and two (1.2%) breaching the terms of their licence (ITT control group: n = 3; 1.8%). The biggest difference between the two groups was in the number of security recalls with 63.5% (n = 108) of the EBM group being recalled for security reasons compared to 37.9% of the ITT-matched control group (n = 66). The actual mean time 'at risk' for the control group (307.08 days) was not significantly longer than the EBM group (282.47 days). Of those released to the community for at least one year, there was no difference in re-offending rates between the EBM and ITT-matched control group (8.8% vs. 6.1%). Those in the EBM group were however more likely to be recalled in their first year of release compared to the ITT-matched control group (29.6% vs. 14.3%). The difference in recall rate was statistically significant χ^2 (2, N = 272) = 11.16, p = .004.

Multivariate Analysis. A sequential logistic regression analysis was performed to determine the predictive ability of group membership (EBM or control) for failure after

controlling for the effects of time at risk. At Block 1, the expected time-at-risk variable was entered followed by the group membership variable at Block 2.

The constant-only model containing the intercept was significant χ^2 (1, N = 342) = 18.36, p < .001. The time-at-risk variable, entered at Block 1, was also significant, χ^2 (1, N = 342) = 10.39, p = .001. The model correctly classified 64.3% of cases but was more effective at correctly identifying failures (95.3%) compared to non-failures (14.5%). The model was not a good fit according to the Hosmer and Lemeshow test, χ^2 (8, N = 342) = 46.39, p < .001. Group membership, entered at Block 2, also predicted failure χ^2 (1, N = 342) = 22.50, p < .001. In comparison to the model at Block 1, there was a marginal increase in the number of cases being correctly classified overall in Block 2 (67.0%), with improved specificity at predicting non-failures (37.4%) albeit at the expense of identifying failures (85.3%). The model fit remained poor, χ^2 (8, N = 342) = 33.57, p < .001. An AUC of .679 indicated moderate model predictive accuracy. Table 2 shows how the predictor variables contributed to the model. The odds ratio (Exp [B]) indicates that assignment to the EBM group increased the odds of failure from a factor of two to a factor of three.

A Kaplan-Meier test showed that assignment to the EBM group reduced the survival time in open conditions compared to the control group χ^2 (2) = 19.11 p < .001. Assignment to the EBM group was associated with an increased hazard of failure by 112% (HR = 2.12; CI = 1.60, 2.80).

Table 2

Intention to Treat Analysis: Logistic Regression of Failure as a Function of Time-at-Risk and Group Membership (Intervention, control)

Block 0					
Variable	*B*	*SE*	*Wald*	*Exp (B)*	*95% CI*
Constant	0.48	0.11	18.36	1.61	
Block 1					
Variable	*B*	*SE*	*Wald*	*Exp (B)*	*95% CI*
Expected time at risk (days)	-0.00	0.00	10.38	1.00	[0.99-1.00]
Constant	1.21	0.25	22.63	0.00	
Block 2					
Variable	*B*	*SE*	*Wald*	*Exp (B)*	*95% CI*
Time at risk (days)	-0.02	0.01	12.61	1.00	[0.99-1.00]
Group membership (Int = 1, Cont = 0)	1.13	0.28	22.50	3.11	[1.95-4.97]
Constant	0.78	0.27	8.43	2.19	

Note. Classification Accuracy Block 0 = 61.7%, Block 1=64.3%, Block 2=67.0%

Treatment Received (TR) Analysis

Univariate Analysis. Of the 171 offenders in the intervention group, 78 (45.6%) were discharged from EBM per-protocol (completers), 71 (41.5%) started but failed in open conditions during the course of EBM (non-completers), and 22 (12.9%) failed in open conditions before EBM commenced (non-starters). The descriptive statistics for the three intervention groups alongside the control group are shown in Table 3. Kruskal-Wallis and chi-square analyses confirmed that there were no significant differences between the four groups on the matching variables: behavioural warnings χ^2

(3, N = 342) = 6.12, p = .106; OVP score χ^2 (3, N = 342) = 4.03, p = .258; or personality disorder diagnosis χ^2 (9, N = 342) = 9.29, p = .411.

Table 3. *Descriptive Statistics by 'Treatment Received' Group*

Variable	**Completers (*n* = 78)**	**Non-completers (*n* = 71)**	**Non-starters (*n* = 22)**	**ITT-matched control group (*n* = 171)**
Age, M_{years} *(SD)*	39.6 *(11.6)*	36.3 *(11.5)*	34.1 *(8.4)*	35.6 *(10.7)*
Previous convictions, M_{number} (*SD*)	14.0 *(11.6)*	*15.9 (11.2)*	15.6 *(14.9)*	13.2 *(10.7)*
Index offence (%)				
Burglary	7.8	9.9	4.5	9.8
Drug	0.0	1.4	4.5	2.4
Fraud	0.0	0.0	0.0	0.0
Robbery	24.7	19.7	31.6	19.5
Sexual	26.0	23.9	4.5	20.1
Violence	32.5	36.6	54.5	41.5
Other	9.0	8.5	0.0	6.7
Sentence type (%)				
Determinate	17.9	23.9	27.3	44.4
Indeterminate for Public Protection	43.6	50.7	45.4	35.7
Life	38.5	25.4	27.3	19.9
Behavioural warnings in 12 months prior to transfer to open conditions M_{number} (*SD*)	3.6 *(4.0)*	4.4 *(4.1)*	4.8 *(3.3)*	3.6 *(3.6)*
Personality Disorder Diagnosis				
Not assessed	41.0	54.9	54.5	39.8
Negative diagnosis	18.0	7.0	9.1	17.0
Subthreshold traits	25.6	26.8	27.3	31.0
Positive diagnosis	15.4	11.3	9.1	12.3
OASys Violence Predictor (OVP) score M_{number} (*SD*)	32.4 *(15.7)*	37.6 *(15.8)*	32.5 *(12.4)*	35.2 *(17.4)*

Thirty-six completers (46.2%) failed in open conditions. This compared to a failure rate of 85 (49.7%) in the ITT-matched control group (n = 171). Forty men failed (51.3%) in the TR-matched control (n = 78), χ^2 (1, N = 154) = 0.23, p = .629. The difference in failure rate between the completer group and the TR-matched sample was not statistically significant and was equivalent across each of the serious recidivistic events with four (5.1%) of the EBM group absconding (TR control group: n = 4; 5.1%), one (1.3%) re-offending (TR control group: n = 0; 0.0%) and two (2.6%) breaching the terms of their licence (TR control group: n = 3; 3.9%). Of those released to the community, ITT-matched control group cases (n = 147) were much less likely to be re-incarcerated for any reason (n = *30;* 20.4%), compared to EBM completers (n = 22/63; 34.9%), non-completers (n = 18/47; 38.3%) and non-starters (n = 8/15; 53.3%). The finding was statistically significant χ^2 (6, N = 272) = 14.44, p = .025. Whilst the community re-offending rates between completers and their TR-matched controls (n = 130) were equivalent (8.1% vs 8.8%), completers were twice as likely to be recalled (27.4% vs. 13.2%) compared to the TR-control group. This was statistically significant χ^2 (2, N = 119) = 4.08, p = .044.

Multivariate Analysis. A sequential logistic regression analysis was performed to determine the predictive ability of group membership (completer or matched control) for failure, after controlling for time at risk. As previously, at Block 1 the expected

time-at-risk variable was entered followed by the group membership variable at Block 2.

Table 4

'Treatment Received 'Analysis: Logistic Regression of Failure as a Function of Time-at-Risk and Group Membership (Completer, Matched Control)

Block 0					
Variable	*B*	*SE*	*Wald*	*Exp (B)*	*95% CI*
Constant	-0.51	0.16	0.10	0.95	
Block 1					
Variable	*B*	*SE*	*Wald*	*Exp (B)*	*95% CI*
Expected time at risk (days)	-0.00	0.00	3.76	1.00	[0.99-1.00]
Constant	0.60	0.37	2.66	1.83	
Block 2					
Variable	*B*	*SE*	*Wald*	*Exp (B)*	*95% CI*
Time at risk (days)	-0.01	0.01	3.47	1.00	[0.99-1.00]
Group membership (Int = 1, Cont = 0)	-0.08	0.33	0.06	0.92	[0.48-1.76]
Constant	0.63	0.38	2.68	1.88	

Note. Classification Accuracy Block 0 = 51.3%, Block 1=61.5%, Block 2=61.5%

Table 4 shows the logistic regression model. The constant-only model containing the intercept was not significant χ^2 (1, N = 156) = 0.10, p = .749. At Block 1, the model, upon entering the expected time-at-risk variable was non-significant, χ^2 (1, N = 156) = 3.76, p = .053. The model correctly classified 61.5% of cases, 64.5% of failures and 58.8% of non-failures, but did not produce a good model fit based on the Hosmer and Lemeshow test, χ^2 (8, N = 156) = 18.24, p = .020. Group membership was added at

Block 2 but did not contribute significantly to the model χ^2 (1, N = 154) = 0.06, p = .804. There was no improvement in classification accuracy with 63.2% of failures and 60.0% of non-failures correctly identified. Overall, EBM completers were no more or less likely to succeed or fail compared to the TR-matched control group.

Discussion

This study aimed to determine whether Enhanced Behaviour Monitoring (EBM) is effective at reducing failures in open prisons in England and Wales. Behavioural monitoring can unlock risk-related information, inaccessible to other risk assessment methodologies, and act as a foundation from which to intervene at critical moments in an offender's post-release journey away from crime. There is precedent in forensic samples for successfully intervening on the basis of behavioural monitoring protocols (Tolisano et al., 2017; Ward & Bosek, 2002). EBM contains a feedback loop to encourage self-management of risk and, on that basis, we hypothesised that EBM would impact positively on 'failure' rates, despite the additional monitoring.

The hypothesis was not supported by the data. Using an Intention To Treat (ITT) approach with matched control group, the results showed, that whilst there were no differences in serious failure outcomes between the intervention and ITT-matched control group, those allocated to EBM were returned to closed conditions sooner, and at greater frequency; indicating perhaps that EBM is primarily being used as a surveillance tool, to detect rule violations (Hyatt & Barnes, 2017; Petersilia & Turner, 1993). Those

completing EBM per-protocol in the Treatment Received (TR) analyses were more likely to survive in open conditions than the TR-matched controls. Yet those completing EBM per-protocol in the Treatment Received (TR) analyses were slightly more likely to survive in open conditions compared to the TR control group, with no increase in serious failure outcomes, albeit the finding did not reach significance. Overall, we were unable to demonstrate the efficacy of behavioural monitoring – namely EBM – as a tool for reducing failures in open conditions by rehabilitative means.

Yet, before unequivocally discrediting the potential of EBM as an intervention, it is important to acknowledge the methodological limitations underpinning the findings. Despite best efforts to match members of the control group to the EBM group, a common criticism of Propensity Score Matching (PSM) analysis is that it may omit, principally due to nonrecognition, the effect of several unmeasured but clinically relevant factors that can affect the outcome (Reiffel, 2020). Our study was limited to matching on the basis of three variables only (i.e., behavioural warnings in custody, OVP score, personality difficulties), as identified in Goodley and Pearson (2023). These data were collected from administrative databases; but propensity scores derived from largely administrative data do not necessarily balance unmeasured clinical confounders (Austin et al., 2005). Indeed, a relatively small proportion of the control group – just one-third – could be matched to the intervention group indicating that the

two groups were already distinct and classifiable at the outset. It is possible therefore that the matching variables may have been acting as a proxy for a particular clinical profile represented more frequently in the EBM group and which, based on administrative data, was matched superficially in the control group. Indeed, the decision to allocate to EBM is decided by psychological professionals trained in risk assessment. It is likely that these professional decisions were based in an aggregation of risk-relevant factors, meaning the potential available control group was comprised of a greater proportion of men who were likely to be successful in open conditions. In this context, the EBM cohort performed surprisingly well and therefore it is possible to argue that the per-protocol results may be considered to represent a conservative estimate (possible under-estimate) of the effect of EBM on failure rates.

Nonetheless, despite the limitations in directly comparing the outcomes of the intervention group against the outcomes for the control group, there existed patterns in the descriptive data which indicated a 'blunting' of any intended rehabilitative impact of EBM. First, the ratio of security recalls to the rate of community re-offending was approximately 7:1 across both the EBM group (n = 108/15*)* and the control group (n = 66/10). That is, for every community re-offence by an individual in each group, an equivalent of seven individuals were returned to closed conditions. Ostensibly therefore, the approach to non-compliance remained consistent regardless of EBM allocation; additional time/space for behavioural improvement was not afforded to those

allocated to EBM. Second, whilst the sample was purposively selected and included those at highest risk of open prison failure, the one-year community re-offending rates were low across both groups (intervention 8.8%; control 6.1%), certainly relative to the security recall rate. Indeed, across both groups, recall was initiated after just one adjudication on average, perhaps indicating low risk tolerance and the inclination to recall based on 'single-episode' behaviours rather than behavioural aggregation; the latter being a much better measure of antisocial persistence (Hanson et al., 2007; McDougall et al., 2013). As argued by Goodley and Pearson (2023), and as per the tenet of this paper, it should be possible to reduce in-prison failure rates without adversely impacting the community recidivism rate.

The intensive supervision literature contains key lessons for EBM. 'Surveillance only' supervision programmes serve only to increase the rate of reincarceration due to technical violation but have null effects on recidivism (Drake, 2018; Hyatt & Barnes, 2017; Petersilia & Turner, 1993). Only those supervisory approaches which incorporate a therapeutic component or where supervisors act as 'agents of change' positively impact recidivism outcomes (Bonta et al., 2011; Drake, 2018; Petersilia & Turner, 1990; Taxman, 2008). A process evaluation is a necessary next step to examine the implementation of EBM to seek ways to reinforce the rehabilitative component. Likewise, the organisational context or philosophy of the institutions in which EBM has been implemented may be important. Clear and Latessa (1993) found that the personal

role preferences of probation officers (i.e., law enforcement vs. social worker) is mediated by the philosophy of the institution. Goodley and Pearson (2023) have previously shown that open prisons in England take a conservative approach to non-compliance – a function of the socio-political context in which these open prisons operate – which potentially permeates the rehabilitative potential within EBM. Muir-Cochrane and colleagues (2012) refer to a phenomenon of 'anxious vigilance' in which staff, fearful of the consequences of serious failure outcomes such as absconding, are hypervigilant to the threat, at the expense of patient care. Conversely, good relational security and a supportive culture may be more promising as risk management strategies – to stabilise behaviour and foster desistance processes (Andvig et al., 2021; Auty & Liebling 2020; Bystrom Borg & Törnell, 2017) – certainly in comparison to regimes with a law enforcement orientation (Alexander, 2006; Auty & Liebling, 2020).

Practical implications

Whilst we have been unable to demonstrate the effectiveness of Enhanced Behaviour Monitoring (EBM) to stabilise behaviour, and so advocate for the use of behavioural monitoring protocols as an intervention in open prisons internationally, several practical points should be noted. First, despite purposively sampling those deemed by practitioner psychologists to have a raised risk profile for failure (i.e., the intervention group), 8.2% absconded or re-offended in custody and 8.8% re-offended within one-year of being placed in the community. Manifestly therefore, the recall rate

of 63.2%, was grossly disproportionate to the risk. Given open prisons' purpose is to prepare prisoners for release thereby reducing community recidivism, on the balance of risk, open prisons should defer recall decisions in favour of allowing a greater accumulation of behavioural data. McDougall and colleagues (2013) have previously shown that the frequency of concerning behaviours in custody and the community were significantly correlated and that, based on that frequency, community recidivism could be predicted with 92% accuracy. Therefore, behavioural schemes such as EBM could be used effectively to inform community release management rather than being a tool for managing risk *while in* open prisons. Goodley and Pearson (2023) argue that systems-level policy implementation should be the primary means of ensuring effective risk management in open prisons. Basing re-categorisation decisions in comprehensive risk assessment, softening transitions to and within open prisons, and building secure accommodation provision into open sites, is the measured approach to risk management; certainly preferable to dropping unsuitable candidates into an unstructured environment and relying on open prisons 'spotting' or 'catching' the supposed residual risk. Indeed, there is no precedent for using behavioural schemes purely as a surveillance tool without a rehabilitative component. Drake (2018), using a Monte Carlo simulation analysis of the likely risk and uncertainty of investing in various Intensive Supervision Programmes (ISPs) based on their meta-analytic effect size for reducing recidivism, showed that the benefits of investing in ISPs with a rehabilitative

component far outweighed the costs. 'Surveillance only' programmes had a null effect on recidivism at best. We argue that open prisons should use behavioural monitoring protocols, not as a tool for defending against risk exposure, but to support and underpin risk formulations to support effective management on release.

Conclusions

The results of the current study examining the effectiveness of a behavioural monitoring scheme (EBM) for reducing failures in open conditions in England and Wales was inconclusive. Despite our best efforts to match the EBM group to a suitable control, there are reasons to believe that the two groups differed from one another at the outset and the, largely administrative-level, variables identified in the literature as relevant to failure in open conditions were insufficient to balance unmeasured clinical confounds. Nonetheless, an examination of the descriptive data indicated that prison managers adopted a similar approach to behavioural aberrations across both groups, indicating, at best, that EBM is having a null impact on behavioural stabilisation. We do not view EBM as a flawed policy rather the flaws lie in the implementation. Reductions in recidivism are possible with behavioural monitoring interventions (Tolisano et al., 2017; Ward & Bosek, 2002) and intensive supervision programmes with a rehabilitative component (e.g., Bonta et al., 2011). A process evaluation is necessary to understand the barriers to realising the rehabilitative potential within EBM.

Consideration should also be given to re-purposing EBM as a tool for formulating and steering risk management on release to the community.

[The author(s) received no financial support for the research, authorship, and/or publication of this article].

[The authors report there are no competing interests to declare].

References

Adams K (1992). Adjusting to prison life. *Crime and Justice*, 16, 275-359.

Alexander J (2006). Patients' feelings about ward nursing regimes and involvement in rule construction. *Journal of Psychiatric and Mental Health Nursing,* 13(5), 543-553.

Andvig E, Koffeld-Hamidane S, Ausland LH & Karlsson B (2021). Inmates' perceptions and experiences of how they were prepared for release from a Norwegian open prison. *Nordic Journal of Criminology*, 22(2), 203-220.

Atkinson DF and Mann RE (2012). Prison officers as observers of offence paralleling behaviours: an untapped resource? *The Journal of Forensic Psychiatry & Psychology*, 23(2), 139-155.

Austin PC (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399-424.

Austin PC, Mamdani MM, Stukel TA, Anderson GM and Tu JV (2005). The use of the propensity score for estimating treatment effects: administrative versus clinical data. *Statistics in Medicine*, 24(10), 1563-1578.

Auty KM & Liebling A (2020). Exploring the relationship between prison social climate and reoffending. *Justice Quarterly*, 37(2), 358-381.

Barnes GC and Hyatt JM (2018). Intensive Supervision Probation. In *Oxford Research Encyclopedia of Criminology and Criminal Justice*. https://doi.org/10.1093/acrefore/9780190264079.013.265

Bonta J, Bourgon G, Rugge T, Scott TL, Yessine AK, Gutierrez L and Li J (2011). An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior,* 38, 1127-1148.

Bottoms AE (1999). Interpersonal violence and social order in prisons. *Crime and Justice,* 26, 205–281.

Byström Borg C & Törnell A (2017). Frihet under ansvar på Gruvbergets öppna anstalt–en kvalitativ studie om vilken betydelse Gruvbergets verksamhet har för de intagna. Retrieved from: https://www.diva-portal.org/smash/get/diva2:1117935/FULLTEXT01.pdf

Clark DA, Fisher MJ and McDougall C. (1993). A new methodology for assessing the level of risk in incarcerated offenders. *The British Journal of Criminology*, *33*(3), 436-448.

Clear TR and Latessa EJ (1993). Probation officers' roles in intensive supervision: Surveillance versus treatment. *Justice Quarterly*, 10(3), 441-462.

Daffern M, Jones L, Howells K, Shine J, Mikton C and Tunbridge V (2007). Refining the definition of offence paralleling behaviour. *Criminal Behavior and Mental Health,* 17*, 265-273.*

Dawar A and Davis M (2014, May 21). Scandal of our 'holiday camp' open prisons. *Daily Express*. https://www.express.co.uk/news/uk/497892/Scandal-of-our-holiday-camp-open-prisons

Dehejia RH and Wahba S (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1), 151-161.

Douglas KS and Skeem JL (2005). Violence risk assessment: getting specific about being dynamic. *Psychology, Public Policy, and Law*, 11(3), 347.

Drake EK (2018). The monetary benefits and costs of community supervision. *Journal of Contemporary Criminal Justice*, 34(1), 47-68.

Eckhouse L, Lum K, Conti-Cook C and Ciccolini J (2019). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, 46(2), 185-209.

Goodley G, Pearson DA and Morris P (2022). Predictors of recidivism following release from custody: a meta-analysis. *Psychology, Crime & Law*, 1-27.

Goodley G & Pearson D (2023). Monitoring prisoners preparing for release: Who 'fails' in open prison conditions?. *European Journal of Criminology*, 14773708231183570.

Gordon, A and Wong, SCP (2010). Offence analogue behaviours as indicators of criminogenic need and treatment progress in custodial settings. In M Daffern, L Jones and J Shine (Eds.), *Offence Paralleling Behaviour: A case formulation approach to offender assessment and intervention* (pp. 171–184). Chichester: Wiley-Blackwell.

Green B, Carroll A and Brett A (2010). Structured risk assessment in community forensic mental health practice. *Australasian Psychiatry*, 18(6), 538-541.

Greenall PV and Richardson C (2015). Adult male-on-female stranger sexual homicide: A descriptive (baseline) study from Great Britain. *Homicide studies*, 19(3), 237-256.

Hanson RK, Harris AJ, Scott TL and Helmus L (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Vol. 5,

No. 6). Public Safety Canada. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/index-en.aspx

Harris GT and Rice ME (2015). Progress in violence risk assessment and communication: Hypothesis versus evidence. *Behavioral Sciences & the Law*, 33(1), 128-145.

Helmus LM and Babchishin KM (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*, 44(1), 8-25.

Ho A, Lesneskie E and Hsu KH (2018). What we don't know may hurt us: an examination of systematic bias in offender risk assessments. *Deviant Behavior*, 39(12), 1566-1577.

Hollin CR (2008). Evaluating offending behaviour programmes: Does only randomization glister?. *Criminology & Criminal Justice*, 8(1), 89-106.

Home Office (2006). *Offender Assessment System OASys user manual.* London: Home Office.

Hyatt JM and Barnes GC (2017). An experimental evaluation of the impact of intensive supervision on the recidivism of high-risk probationers. *Crime & Delinquency*, 63(1), 3-38.

Jones LF (2000). *Identifying and working with clinically relevant offence paralleling behaviour.* Paper presented at Division of Clinical Psychology, Forensic Special Interest Group, Nottinghamshire.

Jones LF (2004). Offence Paralleling Behaviour (OPB) as a Framework for Assessment and Interventions with Offenders. In A Needs and G Towl (Eds.), *Applying Psychology to Forensic Practice* (p. 34–63). Oxford: Blackwell and British Psychological Society.

Kovach CR (2020). Intention to treat: Necessary but not sufficient. *Research in Gerontological Nursing*, 13(2), 62-63.

Lee J and Little TD (2017). A practical guide to propensity score analysis for applied clinical research. *Behaviour Research and Therapy*, 98, 76-90.

McDougall C, Clark DA and Woodward R (1995). Application of operational psychology to assessment of inmates. *Psychology, Crime and Law*, 2(2), 85-99.

McDougall C, Pearson DA, Willoughby H and Bowles, RA (2013). Evaluation of the ADViSOR project: Cross-situational behaviour monitoring of high-risk offenders in prison and the community. *Legal and Criminological Psychology*, 18(2), 205-228.

Ministry of Justice. (2019). *Release on Temporary Licence policy framework.* London: Ministry of Justice.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1011503/rotl-pf.pdf

Ministry of Justice (2020). *Security Classification Policy Framework.* London: Ministry of Justice. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1011502/security-categorisation-pf.pdf

Monahan J, Skeem J and Lowenkamp C (2017). Age, risk assessment, and sanctioning: Overestimating the old, underestimating the young. *Law and Human Behavior*, 41(2), 191-201.

Mooney JL and Daffern M (2015). The relationship between aggressive behaviour in prison and violent offending following release. *Psychology, Crime & Law*, 21(4), 314-329.

Muir-Cochrane E, van der Merwe M, Nijman H, Haglund K, Simpson A and Bowers L (2012). Investigation into the acceptability of door locking to staff, patients, and visitors on acute psychiatric wards. *International Journal of Mental Health Nursing*, 21(1), 41-49.

National Offender Management Service (2015). *Enhanced Behaviour Monitoring: (PSI 24/2015; PI 16/2015)*.

Paparozzi MA and Gendreau P (2005). An intensive supervision program that worked: Service delivery, professional orientation, and organizational supportiveness. *The Prison Journal*, 85(4), 445-466.

Pearson DA and McDougall C (2017). Serious offenders: using evidence to predict and manage the risk. *Behavioral Sciences & the Law*, 35(3), 239-252.

Petersilia J and Turner S (1990, December). *Intensive supervision for high-risk probationers: Findings from three California experiments*. Santa Monica, CA: RAND.

Petersilia J and Turner S (1993). *Evaluating intensive supervision probation/parole: Results of a nationwide experiment*. US Department of Justice, Office of Justice Programs, National Institute of Justice.

Reiffel JA (2020). Propensity score matching: The 'Devil is in the details' where more may be hidden than you know. *The American Journal of Medicine*, 133(2), 178-181.

Rosenbaum PR and Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Rosenbaum PR and Rubin DB (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.

Schmidt S, Bliesener T and van der Meer E (2019). Risk and protective factors of delinquency that are sensitive to migration and culture. *Psychology, Crime & Law*, 25(8), 847-873.

Shepherd SM and Lewis-Fernandez R (2016). Forensic risk assessment and cultural diversity: Contemporary challenges and future directions. *Psychology, Public Policy, and Law*, 22(4), 427-438.

Simpson AI, Penney SR, Fernane S and Wilkie T (2015). The impact of structured decision making on absconding by forensic psychiatric patients: results from an AB design study. *BMC Psychiatry*, 15(1), 103-116.

Soothill K, Francis B, Ackerley E and Fligelstone R (2002). *Murder and serious sexual assault: What criminal histories can reveal about future serious offending* (Police Research Series Paper 144). London: Home Office.

Stuart EA (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: A Review Journal of the Institute of Mathematical Statistics*, 25(1), 1–28.

Taxman FS (2008). No illusions: Offender and organizational change in Maryland's proactive community supervision efforts. *Criminology & Public Policy,* 7, 275-302.

Tolisano P, Sondik TM and Dike CC (2017). A positive behavioral approach for aggression in forensic psychiatric settings. *The Journal of the American Academy of Psychiatry and the Law*, 45(1), 31-39.

Venner S, Sivasubramaniam D, Luebbers S and Shepherd SM (2021). Cross-cultural reliability and rater bias in forensic risk assessment: A review of the literature. *Psychology, Crime & Law*, 27(2), 105-121.

Vojt G, Thomson LD and Marshall LA (2013). The predictive validity of the HCR-20 following clinical implementation: does it work in practice?. *Journal of Forensic Psychiatry & Psychology*, 24(3), 371-385.

Ward KM and Bosek RL (2002). Behavioral risk management: Supporting individuals with developmental disabilities who exhibit inappropriate sexual behaviors. *Research and Practice for Persons with Severe Disabilities*, 27(1), 27-4.