The JABBS Foundation

UNIVERSITY OF BIRMINGHAM

Women, Crime & Justice Research Group

Women's Risk Needs Assessment UK Validation:

Preliminary Findings Report

Pemberton S., Long., J. Holdsworth, E., Mackenzie, K., Ruprai, D., & Summers, R.

Acknowledgements

This research was made possible through Professor Emily Salisbury's continued support and guidance. We thank Noble Software, The Data Linkage Team, and Lauren Small at the Ministry of Justice for their assistance. We're especially grateful to The JABBS Foundation for their continued funding, support, and infrastructure which enabled the validation study and the subsequent Effective Women's Centre Project. This research represents a collaborative, multidisciplinary effort and we thank the frontline caseworkers and professional staff at all participating women's centres for their valuable contributions.

We dedicate this research to Dr. Susie Balderston, who passed away in late 2023. Susie contributed significantly to this work as a Research Fellow in her time at the University of Birmingham from 2017-2021. Her work focused on the harms faced by criminal justice-involved women, domestic violence survivors, sex workers, and victims of disablist hate crime. She led the "Tackling Violence Against Disabled Women & Girls" project at Vision Sense with the Centre for Disability Research. Susie also served as an advisor to the Equality and Human Rights Commission's Statutory Inquiry into Disability Harassment. At Lancaster University's UNESCO Violence & Society Centre, she worked on victim assistance for women trafficked for sexual exploitation for the European Commission. Before joining Birmingham in 2017, she lectured on sexual violence, social policy issues and disability studies at Lancaster, Salford, and Durham Universities. Her colleagues and friends across the sector deeply miss her.

About the Women, Crime & Justice Research Group

Prof Simon Pemberton is currently a Professor in Social Policy and Criminology at the University of Birmingham. He is director of the Women, Crime, and Justice Research Group, and has published widely on aspects of social harm, crime and criminalisation, as well as issues in relation to poverty and social exclusion. He currently is PI for the 'Effective Women's Centre' project which explores the effectiveness of community-based gender responsive programmes for justice-involved women.

Dr. Joanna Long is a Senior Research Fellow on the WRNA project and has diverse experience managing large scale research studies in complex environments (NHS, MoJ and Voluntary sector) using both quantitative and qualitative research methods. She has worked with Professor Simon Pemberton since 2017 and is part of the Women, Crime, and Justice Research Group who are dedicated to transforming the lives of women involved in the criminal justice system by providing the evidence base for 'what works' in women's centres.

Dr. Denise Ruprai is a Research Fellow on the WRNA project, investigating Trauma and Health in justice involved women, with multiple complex needs, such as substance misuse and mental health difficulties. She has with extensive knowledge of neuropsychology-imaging, anatomy, psychiatry- and trauma related research, along with versatile experience in complex data analysis and statistics.

Dr. Kelly Mackenzie is a Research Fellow on the WRNA project. She has both an academic and practice-based background having supported victim-survivors in a domestic abuse service following the completion of her PhD. Kelly is also a member of the organising committee for the Violence Against Women and Girls Research Network (VAWGRN).

Dr. Ella Holdsworth is a Research Fellow on the WRNA project. Ella has also undertaken a number of research posts. These include research associate on a project which explored police responses to child to parent violence (2019-20) and a European comparative project on the use of electronic monitoring on five EU Member States (2015-16).

Dr. Richard Summers is a Research Fellow on the WRNA project. He is currently contributing to the research methods and statistics of the project. He has a background in cognitive neuroscience and population health.

Version: May 2025

Abstract

Abstract

This study presents the first comprehensive validation of the Women's Risk Needs Assessment (WRNA) within the UK context, examining its psychometric properties and predictive validity among justice-involved women accessing community-based services. Using a prospective design, 506 women (mean age 37.3 years) from three women's organisations across seven sites completed WRNA assessments, with data linked to Police National Computer (PNC) records for 12-month follow-up analysis (84% linkage success, n=426).

Predictive validity analyses demonstrated robust stepwise increases in reoffending rates across risk categories: Low-Risk (5.2%), Moderate-Risk (16.1%), Medium-Risk (23.0%), and High-Risk (31.8%), with the High-Risk category showing reoffending rates 6.1 times higher than Low-Risk. Spearman correlation analysis revealed a significant positive relationship between overall WRNA scores and conviction counts ($\rho = .22$, $p < .001$). Pairwise chi-square tests with false discovery rate correction confirmed significant differences between High-Risk and Low-Risk groups ($p = .002$), High-Risk and Moderate-Risk groups ($p_{adj} = .013$), and Low-Risk and Medium-Risk groups ($p_{adj} = .013$).

Zero-inflated negative binomial regression revealed that higher WRNA scores significantly predicted both the probability of reoffending (zero-inflation component: $\beta = -0.040$, $p = 0.003$) and offending frequency among those who reoffended (count component: $\beta = 0.025$, $p = 0.024$). Subscale analysis identified nine significant predictors of conviction counts and eight for binary offending outcomes after multiple comparison adjustment. Criminal History emerged as the strongest predictor ($r = .22/.21$), followed by Substance Abuse measures ($r = .18/.21$), Antisocial Friends ($r = .19$), Employment Financial concerns ($r = .15/.14$), and Anger Hostility ($r = .14$). Self-Efficacy demonstrated significant protective effects in both analyses ($r = -.15/-.14$), confirming the importance of strengths-based factors.

Reliability analysis across 24 subscales revealed some variation: six scales achieved excellent reliability ($\alpha/\omega > 0.90$), six demonstrated good reliability (0.80-0.89), six showed acceptable reliability (0.70-0.79), whilst five exhibited questionable reliability (<0.70). The findings confirm the WRNA's utility as a gender-responsive assessment tool with robust dual predictive capacity whilst highlighting specific domains requiring refinement. These results establish an empirical foundation for implementing the WNRA within the British Criminal Justice System.

Contents

Introduction

The overrepresentation of women in the British Criminal Justice System (CJS) with histories of trauma and complex needs has prompted increasing recognition of the necessity for gender-responsive assessment (Corston, 2007; Ministry of Justice, 2018). Traditional criminogenic risk assessment instruments, predominantly developed and validated on male populations, often fail to capture the multifaceted pathways that lead women into criminal justice involvement (Blanchette & Brown, 2006; Gelsthorpe et al., 2007). In response to these limitations, the Women's Risk Needs Assessment (WRNA) was developed as a gender-responsive instrument that identifies the role of gender specific needs that lead to offending, as well as strengths that serve as protective factors. (Van Voorhis et al., 2010).

The assessment is heavily influenced by the pathways perspective, which identifies the distinct trajectories that women and girls take into offending, including childhood victimisation, substance abuse, economic marginalisation, and relational difficulties (Brennan et al., 2012). This perspective recognises that women's pathways into criminal behaviour frequently originate from experiences of trauma, victimisation, and structural disadvantage, factors that are often overlooked by gender-neutral assessment approaches (Bloom et al., 2003; Salisbury & Van Voorhis, 2009). The WRNA aligns with this theoretical orientation by incorporating domains that capture these gendered experiences which include relationship quality, parental stress, mental health, victimisation history, and self-efficacy, alongside traditional risk factors (Van Voorhis, 2012).

Moreover, the WRNA represents a paradigmatic shift from deficit-focused risk assessment to a strengths-based, holistic casework approach. Unlike traditional risk assessment instruments that predominantly measure criminogenic needs, the WRNA deliberately incorporates protective factors and personal strengths that may mitigate risk and support desistance (Blanchette & Brown, 2006; Van Voorhis et al., 2010). This dual focus allows practitioners to develop intervention plans that not only address risk factors but also leverage existing strengths and resources—a particularly salient approach for women, whose offending is often contextualised within relational and socioeconomic challenges (Covington & Bloom, 2007; Wright et al., 2012). The WRNA also calculates a criminogenic risk score which are scored into Low-Risk, Moderate-Risk, Medium-Risk and High-Risk. The risk-banding approach provides a standardised framework for translating complex, multidimensional assessment data into actionable classification decisions.

The validation of assessment tools within specific cultural and institutional contexts is fundamental to ensuring their efficacy and appropriateness (Singh et al., 2011). Whilst the WRNA has demonstrated predictive validity in North American jurisdictions, as well as countries such as the Czech Republic (Smith et al., 2009; Van Voorhis et al., 2010), its applicability and performance within the British CJS, particularly within community-based women's centres, to date has not been examined. This study addresses this critical gap by examining the psychometric properties and predictive validity of the WRNA in a sample of justice-involved women accessing services at women's centres throughout England.

The WRNA has been informed by a significant volume of research, meaning that a strength lies in its demonstrated predictive validity regarding female offender's recidivism and other offending behaviours. Research evidence indicates that the instrument possesses validity for the classification of adult women offenders across both institutional and community corrections settings. The assessment domains have all demonstrated statistically significant correlations with outcome measures including return to prison, technical violations, new arrests, and new convictions. The reliability and validity of assessment instruments are foundational to their ethical and effective implementation. Assessments with poor psychometric properties may lead to misclassification, inappropriate intervention allocation, and ultimately compromised outcomes (Singh et al., 2011). For women, whose pathways to offending and desistance often differ substantially from men's, employing validated gender-responsive tools is particularly important (Blanchette & Brown, 2006). The WRNA's comprehensive structure, encompassing 26 subscales that assess various dimensions of risk, need, and strength, necessitates rigorous evaluation to ensure its utility in guiding effective case management and intervention planning within the British context.

This preliminary report examines the psychometric properties, predictive validity, and practical utility of the WRNA among women accessing services at women's centres across the UK. The validation tested the reliability of WRNA scales and the predictive validity of the WRNA with reoffending outcomes, with the intention of informing the next iteration of the probation WRNA Version 8 and contributing to the broader development of gender-responsive assessment and practice. The findings hold significant implications for policy and practice, potentially informing the development of more effective, tailored approaches to supporting women's desistance from offending and addressing their complex needs within community-based settings.

Methods

Design

A cross-sectional, unrelated design sought to test the internal and predictive validity of the WRNA tool. 506 women completed WRNAs across a 4-6 month period and these scores/risk bandings were then compared to their subsequent criminal justice outcomes across a 12 month follow up period to test the accuracy of the tool. The study design sought to address the following objectives:

Research Objective 1: To understand the ways the WRNA operates and measures risks/needs/responsivities across specific groups of criminal justice involved women and offence types in England.

Research Objective 2: To understand the criminogenic risks/needs/responsivities across the cohort of criminal justice involved women who access women's centres in England.

Research Objective 3: To measure the predictive validity of the first Anglicised version of the WRNA Probation Module v7, checking the weightings and risk scorings provided by the tool for statistical significance and accuracy with criminal justice involved women in England

Participants

Materials

The Women's Risk Needs Assessment (WRNA)

The Women's Risk/Needs Assessment Version 7 (WRNA V7) is a structured clinical interview tool specifically designed for female offenders in probation settings. This validated instrument represents the final validation version adapted for England (May 2021) and serves dual purposes of risk classification for supervision decisions and comprehensive needs assessment for service planning.

The WRNA was initially developed through a collaborative partnership between the National Institute of Corrections and the University of Cincinnati in the United States. Prior to implementation in the current project (2020), Dr Emily Salisbury and the Women, Crime & Justice Research Group (based at the University of Birmingham) (SB, SP, JL) undertook a rigorous process to anglicise the WRNA, ensuring its contextual appropriateness and cultural

relevance for application within the UK criminal justice system. This adaptation process was essential to maintain the assessment's validity whilst accommodating differences in service provision, terminology, and sociocultural context between the North American and British systems.

The WRNA comprises 18 distinct sections with 192 items that assess both static and dynamic risk factors alongside protective strengths. The assessment begins with an exploration of the client's attitudes towards their current offence through open-ended questions, followed by systematic evaluation across multiple criminogenic domains. Core risk assessment areas include criminal history, educational background, employment and financial stability, housing safety, antisocial peer associations, anger and hostility, mental health history and current symptoms, trauma and abuse experiences, substance abuse patterns, gambling behaviours, and family relationships.

Scoring - The WRNA v.7 uses a three-part scoring methodology to generate risk classifications and treatment recommendations. Part 1 assesses eleven risk domains with scores ranging from 0 to 67 points, encompassing both static factors (criminal history, substance abuse history) and dynamic factors (employment/financial status, housing safety, antisocial peers, anger/hostility, current substance abuse, child abuse, adult abuse). Certain domains utilise converted scoring protocols: current depression symptoms are scored as 0 (raw score 0), 1 (raw scores 1-4), or 2 (raw scores 5-6), whilst parental stress employs thresholds of 0 (raw scores 0-9), 1 (raw scores 10-18), or 2 (raw scores 19+).

Part 2 evaluates protective factors across three domains totalling 0 to 7 points: educational strengths (0-4), family support (converted from raw scores where 0=0, 1-3=1, 4=2), and self-efficacy (binary scoring where 0-23=0, 24+=1). The final risk score subtracts Part II from Part I, reflecting the theoretical premise that strengths mediate risk factors.

Risk classifications utilise established cut-offs: Low Risk (≤9), Moderate-Risk (10-21), Medium-Risk (22-36), and High-Risk (≥37). However, these thresholds derive from specific research sites and require recalibration through pilot studies with minimum 12-month follow-up for population-specific validity. Part 4 provides supplementary case-management information for treatment planning but does not contribute to risk classification calculations.

Procedure

This research involved collecting WRNA data at women's centres in England, and linking this data to CJS data through the Ministry of Justice. It involved extensive site set-ups and careful procedural adherence for not only consistency of the data, but for the WRNA to be adopted and utilised in the specific, trauma-informed way that it was intended.

Site Set-up - While the WRNA is in the public domain, it is also proprietary and comes with specific usage conditions, a license agreement, and associated training for its implementation. This ensures that the WRNA is utilised as a trauma-informed tool, reflecting the manner in which it was developed. Before recruitment, all of the fieldwork sites were visited by the research team, or, if this was not possible due to COVID-19 pandemic lockdown restrictions, a series of online site set up meetings were conducted to ensure that there was a thorough understanding by all staff of what the study entailed and how the study would be implemented into standard casework practice. All caseworkers received accredited WRNA end user training, and caseworkers were trained online to identify and consent eligible participants and informed of withdrawal procedures should a participant choose to withdraw from the study.

As a multi-site study, consistency was achieved through training for each caseworker, placement of a researcher from the Research Team at each centre (RP, Anawim; EH, Together Women; KM, Nelson Trust) during the data collection phase, and development of a handbook which was accessible to each caseworker. The handbook was updated through weekly meetings, where the Women, Crime and Justice Research Group at the University of Birmingham (SP, JL, RP, RS, KM, EH) discussed case-by-case complexities and resolved common issues in the assessment questions/scoring to specific instances, to ensure consistent application of the tool across the fieldwork sites.

Recruitment and Consent Procedures - Following their initial attendance at a Women's Centre, potential participants were assessed for study eligibility. During their first routine appointment with an assigned caseworker, eligible women were invited to participate in the research study. Participants could provide consent immediately or request additional time for consideration (up to one month), with subsequent consent obtained by either their caseworker or a research assistant. All participants were provided with a copy of the participant information sheet.

Consent procedures were primarily conducted at the participant's local Women's Centre, though alternative locations including outreach sites or partner organisation premises were used when necessary. The WRNA consent form was completed collaboratively by both participant and caseworker. In cases where consent forms were completed prior to assessment, verbal confirmation of consent was obtained at the commencement of the session.

Data collection was undertaken between July 2021 and May 2023 at the women's centres. The baseline WRNA assessment was conducted during the caseworker/service user first meeting following consent. Subsequently, updates were made to assessments when material changes to life circumstances occurred or when new disclosures emerged during ongoing casework sessions. A final review WRNA was completed four-to-six-months after the baseline assessment. This approach was designed to capture an accurate and comprehensive picture of participants' needs, strengths, and risk factors, with the final assessment score utilised in the predictive validity analysis.

Assessment Administration - The WRNA was administered through a structured interview process. To enhance efficiency, some caseworkers prepared the demographic information page before the appointment and verified these details with participants during the session, while others completed all sections collaboratively with participants. Caseworkers systematically progressed through each section of the assessment, reading questions aloud and adhering closely to the standardised wording. Some caseworkers incorporated support planning during the assessment session, while others deferred this to subsequent appointments. Upon completion of the WRNA, caseworkers finalised centre-specific documentation, such as client agreements and outreach arrangements.

The assessment format facilitated the collection of both structured responses and supplementary case management notes. While all structured components were consistently completed, the volume of free-text documentation varied according to individual case requirements. These qualitative notes, although valuable for casework purposes, were not included in the current analysis.

Data Processing and Quality Control - The WRNA was initially completed in paper format before being uploaded to the participant's digital case plan on the organisation's case management system. Paper copies were subsequently inputted by the research team into Noble (the secure research database) for extraction and analysis. A comprehensive multi-stage quality control protocol was implemented to ensure data integrity and completeness.

The quality assurance process involved several sequential checks: following completion by the caseworker, each WRNA was reviewed by a research assistant (DR, KM, EH) who identified any missing data or scoring errors. When deficiencies were detected, assessments were returned to the original caseworker for clarification or completion before data entry proceeded. Once inputted into Noble, a senior researcher (JL) conducted verification checks by comparing scanned copies of the original paper assessments against the digital entries. Any discrepancies identified during this validation process generated further queries to either the research assistant or caseworker to ensure accurate data capture. This rigorous multi-stage approach was designed to maximise data quality and completeness across all assessment domains. Weekly monitoring meetings supplemented these individual case checks, providing opportunities to address systematic issues and maintain consistency in data collection procedures across all participating sites.

Recognising the potentially sensitive nature of the assessment content and its comprehensive scope, participants were offered the option to complete the WRNA across multiple sessions. This accommodation was implemented either due to time constraints or when deemed appropriate by the caseworker and participant in instances where the assessment addressed traumatic experiences.

Data Analysis Procedure

Data Linkage & Preprocessing

After the WRNA data had been collected, and the cohort completed, Personal Identifiers for the cohort were sent to the Ministry of Justice Data Linkage Team, who linked them to convictions records in the Police National Computer (PNC), removing duplicates, and non-complete matches. The data was successfully matched for 426 participants, representing an 84% success rate.

The reliability analysis utilised the whole dataset of N=506, as this analysis did not require linkage. The predictive validity was assessed by using the linked dataset, consisting of 426 participants.

The Data Linkage Team returned the anonymised PNC dataset that was used for analysis. Once received, it was filtered for records between the WRNA completion date and follow-up date for each participant. (12 months from the WRNA completion). A high proportion of

offences were associated with multiple rows due to multiple disposals. To avoid overcounting, each offence was filtered where disposal rank = 1 or null.

In line with standard practice, for conviction count outcome variables, statistical outliers were identified using the interquartile range (IQR) method and removed where values exceeded 1.5 times the IQR beyond the upper and lower quartiles prior to analysis.

Reliability Analysis

Internal reliability was evaluated for all the WRNA subscales using the whole N=506 dataset. N/A WRNA values were replaced with zero values to maximise available data for analysis. All records were complete for all participants, as this was a requirement for case management.

For scales with three or more items, McDonald's omega (ω) was calculated as the primary reliability coefficient owing to its fewer assumptions regarding tau-equivalence compared to Cronbach's alpha. For scales with fewer than three items, Cronbach's alpha (α) was utilised as an alternative reliability measure. Both coefficients range from 0 to 1, with higher values indicating greater internal consistency.

All analyses were conducted using R (version 4.x.x) with the psych package (Revelle, 2023) for reliability calculations. The *{omega}* function was employed for McDonald's omega calculations using the default settings, whilst the *{alpha}* function was used for Cronbach's alpha. Reliability coefficients ≥0.90 were considered excellent, 0.80-0.89 good, 0.70-0.79 acceptable, 0.60-0.69 questionable and <0.60 poor, following conventional interpretative guidelines (George & Mallery, 2003).

Predictive Analysis

The predictive validity of the WRNA was assessed by linking the data from the Ministry of Justice. Records were pre-processed (see data linkage section). The analytical approach examined predictive validity through three methods: categorical risk analysis, subscale correlation analysis, and regression modelling. Reoffending rates across the four WRNA risk categories were compared using pairwise chi-square tests, with p-values adjusted using false discovery rate correction to control for multiple comparisons. To identify the most predictive domains, correlations were calculated between individual WRNA subscales and two recidivism measures: conviction count (continuous) and any offending (binary). Given violations of normality assumptions, Spearman's rank correlations were employed for continuous outcomes, whilst point-biserial correlations examined binary outcomes. The

Benjamini-Hochberg procedure controlled for multiple comparisons across 33 predictor variables.

A zero-inflated negative binomial (ZINB) regression model examined the relationship between overall WRNA scores and offending frequency. This approach was selected due to the excess of zero counts (most participants did not reoffend) and substantial overdispersion in the data (dispersion ratio = 7.69). The ZINB model simultaneously fits two components; a zero-inflation component predicting the probability of structural zeros (those who would not offend regardless of follow-up duration) and a count component predicting the expected number of offences for those at risk of offending. Model diagnostics confirmed adequate fit through examination of residual plots and predicted versus observed values, whilst the appropriateness of the negative binomial distribution over Poisson was verified through assessment of the dispersion parameter.

To identify the most predictive domains within the WRNA and examine relationships between individual subscales and offending outcomes, correlations were calculated between all WRNA subscales and two recidivism measures: conviction count (continuous) and any offending (binary). This dual analytical approach enabled assessment of which subscales best predict both the likelihood of reoffending and the frequency of offending behaviour among participants.

Given violations of normality assumptions, Spearman's rank correlations were employed for continuous outcomes, whilst point-biserial correlations examined binary outcomes. The Benjamini-Hochberg procedure was applied to control for multiple comparisons across all predictor variables. Where items had been rescored or scaled as part of the standard WRNA risk-scoring protocol, both original and rescored versions were included in the analysis to ensure comprehensive examination of all available subscale variations.

Results

Demographics Table for the Whole Cohort

Variable	Frequency	Percent
Current Violent Offence	215	42.5%
Prior Convictions		
0	248	49%
1-2	107	21.1%
≥3	151	29.8%
Prior Violent Offences	154	30.4%
Prior probation/parole	187	37%
Previously Recalled/Breached/Revoked	77	15.2%
Prior Prison Sentences		
0	398	78.7%
1	47	9.3%
≥3	61	12.1%
Gender		
Female / Woman	500	98.8%
Non-binary / Gender Variant	1	0.2%
Prefer not to say	1	0.2%
Trans	3	0.6%
Trans Woman	1	0.2%
Relationship Status		
Divorced / CP dissolved	15	3.0%
Living with a partner	52	10.3%
Married / Civil Partnership (CP)	21	4.2%
Separate but still legally married/CP	12	2.4%
Single	401	79.2%
Unknown	4	0.8%
Widowed /surviving civil partner	1	0.2%
Disability		
Hearing	12	2.4%
Visual	6	1.2%
Learning	76	15%
Mental Health	341	67.4%
Autism	15	3%
ADHD	36	7.1%
Neurological	3	0.6%
Mobility	44	8.7%
Ethnicity		
Asian or Asian British	14	2.8%
Black/Black British/Caribbean/African	23	4.5%
Mixed or multiple ethnic groups	26	5.1%

Variable	Frequency	Percent
Other Dual	5	1.0%
Unknown	9	1.8%
White	429	84.8%
Religion		
Buddhist	5	1.0%
Christian	155	30.6%
Hindu	1	0.2%
Jewish	1	0.2%
Muslim	27	5.3%
No religion	288	56.9%
Spiritual	16	3.2%
Unknown	13	2.6%
Sexual Orientation		
Asexual	3	0.6%
Don't know	10	2.0%
Gay	8	1.6%
Heterosexual	429	84.8%
Lesbian	25	4.9%
Not Disclosed	12	2.4%
Queer	19	3.8%
Currently Employed		
Full Time	55	10.9%
Part-time/unable	254	50.2%
Unemployed/able	197	38.9%
Single Parent		
No	112	22.1%
Yes	133	26.3%
N/A	261	51.6%
Free School Meals		
No	71	14%
Yes	129	25.5%
N/A	306	60.5%
Registered with GP		
No	59	11.7%
Yes	445	87.9%
Currently in Supported Living	75	14.8%
Renting Private Landlord	75	14.8%
Receiving Food Bank Vouchers	19	3.8%

Predictive Validity

Reoffending rates (Figure 1) demonstrate a robust stepwise increase across risk categories. The Low-Risk group exhibited the lowest reoffending rate at 5.2%, followed by Moderate-Risk (16.1%), Medium-Risk (23.0%), and High-Risk (31.8%). The High-Risk category shows a reoffending rate approximately 6.1 times higher than the Low-Risk category.

The mean sentence counts among those who reoffended (Figure 2) demonstrate a consistent progressive pattern across risk categories. The Low-Risk group showed a median sentence count of 1, Moderate-Risk and Medium-Risk groups both exhibited median sentence counts of approximately 2.3, whilst the High-Risk group demonstrated the highest median sentence count of 3.5. The interquartile ranges, represented by error bars, reveal considerable variability within each risk category, with the High-Risk category showing the greatest spread of sentence counts, suggesting heterogeneity in offending patterns among high-risk individuals who do reoffend.

Figure 1

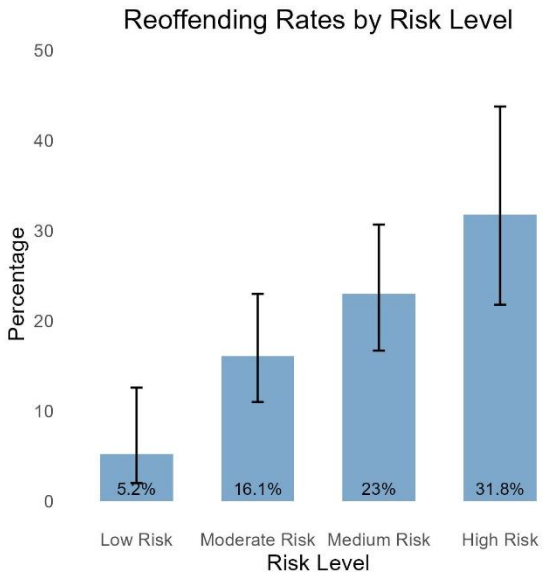
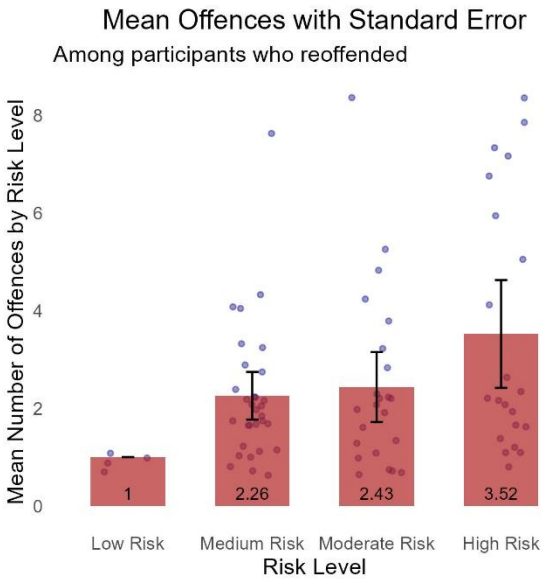


Figure 2



A series of pairwise chi-square tests was conducted to examine the relationship between risk level and offence rates. The p-values were adjusted using the false discovery rate correction method to control for multiple comparisons.

Table 1: *Pairwise Chi-Square Tests Comparing Conviction Rates Across Risk Levels*

Comparison	χ^2	*df*	*p*	p_{adj}
High-Risk vs. Low Risk	11.90	1	< .001	.002
High-Risk vs. Medium-Risk	1.54	1	.158	.158
High-Risk vs. Moderate-Risk	6.64	1	.006	.013
Low Risk vs. Medium-Risk	6.47	1	.006	.013
Low Risk vs. Moderate-Risk	1.78	1	.124	.149
Medium-Risk vs. Moderate-Risk	1.99	1	.118	.149

Table 1 shows significant differences in offence rates were found between High-Risk and Low-Significant differences in offence rates were found between High-Risk and Low-Risk groups (p_{adj} = .002), High-Risk and Moderate-Risk groups (p_{adj} = .013), and Low-Risk and Medium-Risk groups (p_{adj} = .013). No significant differences were observed between High-Risk and Medium-Risk groups (p_{adj} = .158), Low-Risk and Moderate-Risk groups (p_{adj} = .149), or Medium-Risk and Moderate-Risk groups (p_{adj} = .149).

The relationship between overall WRNA score and conviction count was examined using Spearman's rank correlation, revealing a significant positive correlation between the variables (ρ = .22, p < .001). This correlation, though modest in magnitude, provides important evidence that the WRNA maintains predictive validity as a continuous measure, with higher risk scores associated with increased conviction frequency.

Despite the risk-banding showing some inconsistencies in predicting offence frequency, particularly given the large error margins in the low and high-risk groups, the relationship between overall score and conviction count was examined using Spearman's rank correlation. There was a significant positive correlation between the variables, ρ = .22, p < .001. This correlation, though modest in magnitude, provides important evidence that the WRNA maintains predictive validity as a continuous measure, not just when scores are categorised into risk bands. This finding suggests that the instrument captures meaningful variance in offending risk, even if the relationship between risk scores and offence frequency is not perfectly linear or consistent across all risk levels.

Zero-inflated negative binomial Regression

A zero-inflated negative binomial (ZINB) regression model was applied to examine the relationship between WRNA risk scores and conviction count. The model was selected due to

the excess of zero counts (most women did not reoffend, as shown in Figure 3) and evidence of overdispersion in the data. The ZINB model fits two components simultaneously: a count component predicting the number of sentences among those who offend, and a zero-inflation component predicting the probability of structural zeros (those who do not offend at all).

The model diagnostics (Figure 4) demonstrate adequate fit, with residuals showing expected patterns for zero-inflated data. The actual versus predicted plot reveals the model's ability to account for both the excess zeros and the positive counts, though with some underprediction of higher conviction counts.

Figure 3

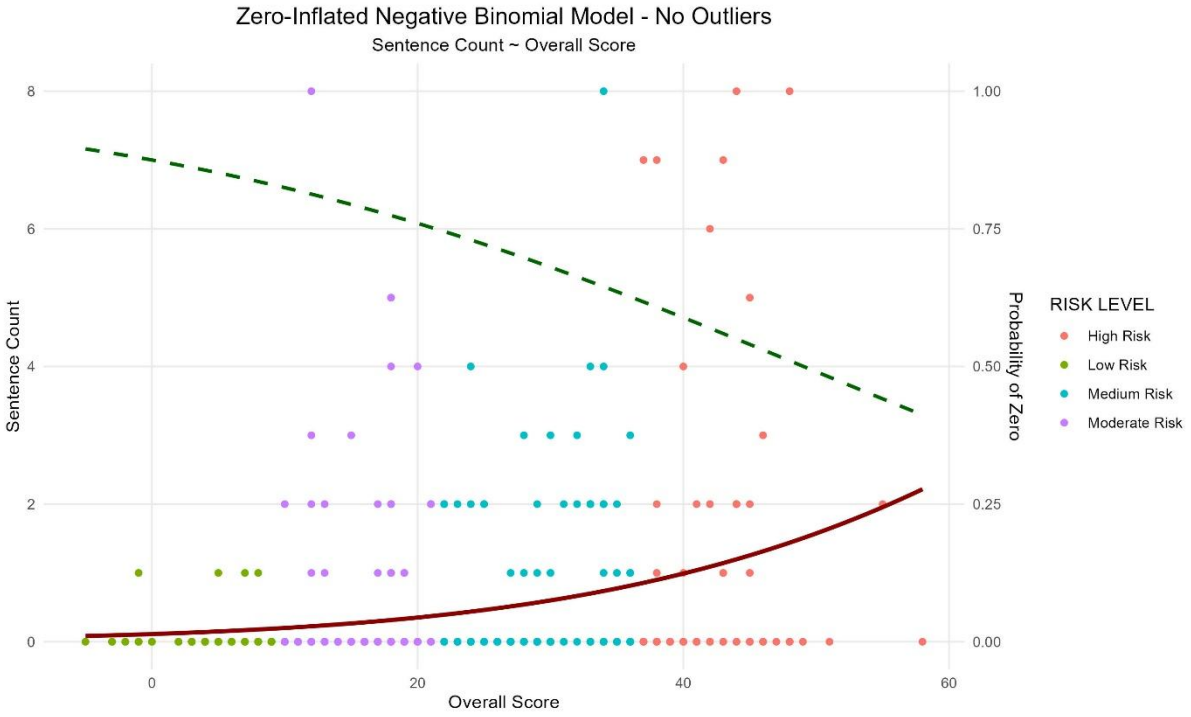
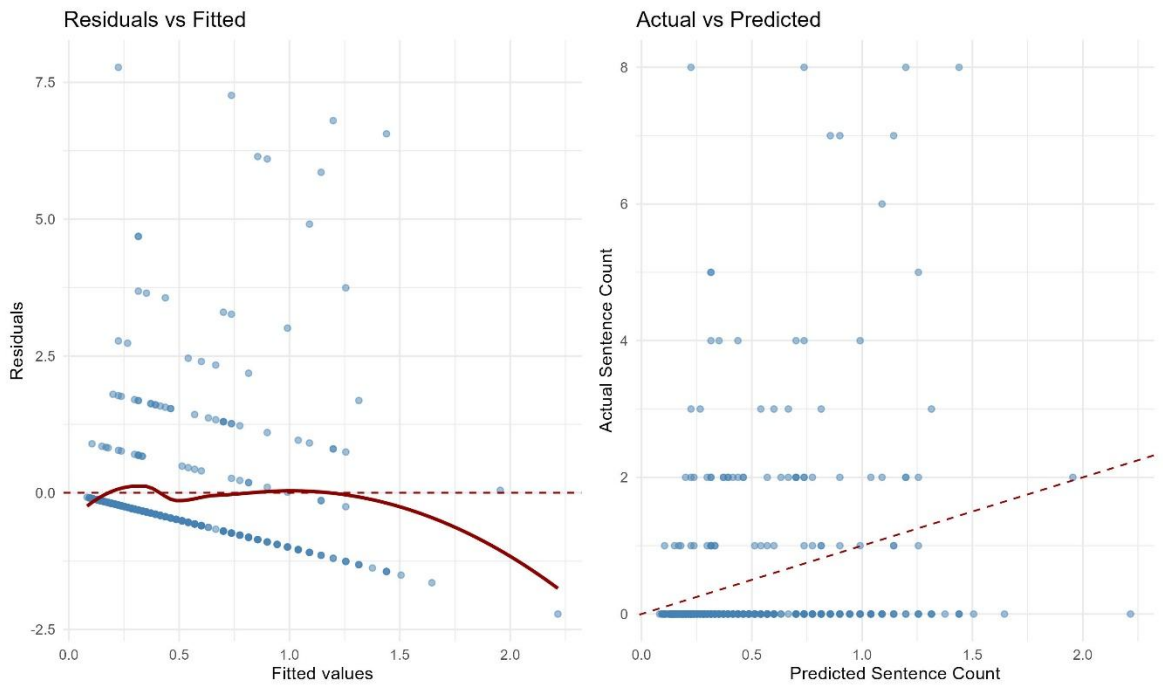


Figure 4



Zero-Inflation Component

The zero-inflation component of the model revealed a significant negative association between WRNA scores and the probability of zero sentences (β = -0.040, SE = 0.013, p = 0.003). This indicates that individuals with higher risk scores were significantly less likely to have zero offences within the 12-month follow-up period. As illustrated in Figure 3 (green dashed line), the probability of remaining offence-free decreases markedly as WRNA risk scores increase. The significant intercept (β = 1.946, SE = 0.427, p < 0.001) indicates a baseline tendency toward structural zeros in the model.

Count Component

The count component of the model, which predicts the number of sentences conditional on offending, showed a significant positive relationship with WRNA scores (β = 0.025, SE = 0.011, p = 0.024). This finding indicates that among women who do reoffend, higher WRNA scores predict greater offending frequency. The non-significant intercept (β = -0.115, SE = 0.386, p = 0.765) suggests no baseline effect when WRNA scores are zero. The dispersion parameter (θ = 1.913) confirms the appropriateness of the negative binomial distribution over a Poisson distribution, indicating notable but manageable overdispersion in the data.

These results demonstrate that the WRNA score functions as a robust predictor across both components of the offending process: it significantly predicts both whether an individual will offend within the 12-month follow-up period and, among those who do offend, the frequency

of their offending behaviour. This dual predictive capacity enhances the clinical utility of the instrument, as it provides meaningful information for both initial risk classification and intensity of intervention planning.

Model Properties

The model diagnostics (Figure 4) demonstrate adequate fit for the zero-inflated negative binomial regression, with residuals displaying expected patterns characteristic of count data with excess zeros. The residuals versus fitted values plot reveals some systematic deviation from linearity, indicated by curvature in the LOESS smoothing line, suggesting the model does not capture all underlying variance in the data. However, this is not uncommon in criminological applications given the inherent complexity of predicting human behaviour.

The actual versus predicted values plot demonstrates the model's capacity to appropriately handle the substantial proportion of participants with zero offences, with most zero predictions aligning accurately with observed outcomes. However, the model exhibits systematic underprediction of higher conviction counts, particularly for individuals with four or more offences during the follow-up period. This conservative bias in predicting high-frequency offending is reflected in the concentration of data points above the diagonal line at higher conviction levels.

Despite these limitations, the model diagnostics confirm the appropriateness of the zero-inflated negative binomial approach for these data. The observed patterns are consistent with established challenges in predicting recidivism outcomes and do not undermine the validity of the substantive findings regarding the WRNA's predictive capacity.

Correlations Between the WRNA Subscales and Number of Convictions

The Spearman correlation analysis identified nine subscales with statistically significant associations with conviction counts after adjusting for multiple comparisons using the Benjamini-Hochberg procedure. This correction method was selected over the more conservative Bonferroni adjustment to balance Type I error control with statistical power, given the exploratory nature of subscale analysis and the risk of excessive Type II errors when examining 33 predictor variables.

Criminal History demonstrated the strongest relationship (r = .22, p < .001), followed by Substance Abuse measures showing consistently robust associations: Total Substance Abuse (r = .21, p < .001), Recent Substance Abuse (r = .21, p < .001), and Historical Substance Abuse (r = .18, p = .001). Antisocial Friends also exhibited a strong positive correlation (r = .19, p

< .001). Additional significant predictors included the Employment/Financial Scale (r = .15, p = .011) and Anger/Hostility Scale (r = .14, p = .015).

Self-Efficacy emerged as a significant protective factor, demonstrating negative correlations in both its original (r = -.15, p = .008) and rescored forms (r = -.13, p = .030), suggesting that higher self-efficacy was associated with fewer convictions. The remaining 24 subscales showed no significant relationships with conviction frequency after adjustment for multiple comparisons.

Table 2 *Correlations Between Sum Variables and Conviction Count*

Predictor	***r***	***p***	p_{adj}
Criminal History Scale	.22***	< .001	< .001***
Substance Abuse (Total)	.21***	< .001	< .001***
Substance Abuse (Recent)	.21***	< .001	< .001***
Antisocial Friends	.19***	< .001	< .001***
Substance Abuse (Historical)	.18***	< .001	.001**
Self Efficacy	-.15**	0.001	.008**
Employment Financial Scale	.15**	0.002	.011*
Anger Hostility Scale	.14**	0.004	.015*
Self Efficacy (rescore)	-.13**	0.008	.030*
Adult Physical Abuse	.10*	0.041	0.117
Parenting Stress	-0.1*	0.042	0.117
Parental Involvement	-0.1*	0.043	0.117
Adult Abuse	.10*	0.049	0.125
Housing Safety Scale	.09	0.056	0.131
Parenting Stress (rescore)	-.09	0.078	0.165
Mental Health - History	.09	0.08	0.165
Childhood Physical Abuse	-.07	0.173	0.318
PTSD	.07	0.178	0.318
Adult Sexual Abuse	.06	0.188	0.318
Attitude Scale	.06	0.193	0.318
Mental Health - Anxiety Depression (rescore)	.06	0.228	0.358
Educational Weaknesses	.05	0.264	0.397
Mental Health - Anxiety Depression	.05	0.346	0.496
Abuse	.04	0.413	0.567
Family Conflict	-.03	0.532	0.702
Childhood Abuse	-.03	0.553	0.702
Relationship Scale	.02	0.655	0.773
Childhood Sexual Abuse	.02	0.656	0.773
Family Support	.02	0.682	0.776
Gambling	-.01	0.821	0.882
Educational Strengths	.01	0.829	0.882
Mental Health - Psychosis	.01	0.912	0.941

Predictor	*r*	*p*	p_{adj}
Family Support (rescore)	.00	0.962	0.962

* $p < .05$, ** $p < .01$, *** $p < .001$. Adjusted p-values calculated using Benjamini-Hochberg method. Spearman's rank correlation was used as normality assumptions were violated for all 33 correlations.

The Spearman correlation analysis identified nine subscales with statistically significant associations with conviction counts after adjusting for multiple comparisons using the Benjamini-Hochberg procedure. This correction method was selected over the more conservative Bonferroni adjustment to balance Type I error control with statistical power, given the exploratory nature of subscale analysis and the risk of excessive Type II errors when examining 33 predictor variables.

Criminal History demonstrated the strongest relationship ($r = .22$, $p < .001$), followed by Substance Abuse measures showing consistently robust associations: Total Substance Abuse ($r = .21$, $p < .001$), Recent Substance Abuse ($r = .21$, $p < .001$), and Historical Substance Abuse ($r = .18$, $p = .001$). Antisocial Friends also exhibited a strong positive correlation ($r = .19$, $p < .001$). Additional significant predictors included the Employment/Financial Scale ($r = .15$, $p = .011$) and Anger/Hostility Scale ($r = .14$, $p = .015$).

Self-Efficacy emerged as a significant protective factor, demonstrating negative correlations in both its original ($r = -.15$, $p = .008$) and rescored forms ($r = -.13$, $p = .030$), suggesting that higher self-efficacy was associated with fewer convictions. The remaining 24 subscales showed no significant relationships with conviction frequency after adjustment for multiple comparisons.

Point-Biserial Correlations between the WRNA Subscales and Offending

The point-biserial correlation analysis examining binary offending outcomes revealed eight subscales with significant associations after adjustment. Criminal History demonstrated the strongest relationship ($r = .21$, $p < .001$), followed closely by Substance Abuse (Total) and Substance Abuse (Recent) (both $r = .20$, $p < .001$), and Antisocial Friends ($r = .19$, $p < .001$). Additional significant predictors included Substance Abuse (Historical) ($r = .18$, $p = .002$), Employment/Financial Scale ($r = .14$, $p = .017$), and Anger/Hostility Scale ($r = .14$, $p = .021$).

Notably, Self-Efficacy emerged as a significant protective factor ($r = -.14$, $p = .021$), indicating that higher self-efficacy was associated with reduced likelihood of any offending during the follow-up period.

Table 3 *Correlations Between Sum Variables and Offended Binary Outcome*

Predictor	***r***	***p***	p_{adj}
Criminal History Scale	.21***†	< .001	< .001***
Substance Abuse (Total)	.20***†	< .001	< .001***
Substance Abuse (Recent)	.20***†	< .001	< .001***
Antisocial Friends	.19***†	< .001	< .001***
Substance Abuse (Historical)	.18***†	< .001	.002**
Employment Financial Scale	.14**†	0.003	.017*
Self Efficacy	-.14**†	0.005	.021*
Anger Hostility Scale	.14**†	0.005	.021*
Self Efficacy (rescore)	-0.12*†	0.016	0.057
Parenting Stress	-0.1*†	0.044	0.146
Adult Physical Abuse	.09†	0.057	0.165
Housing Safety Scale	.09†	0.06	0.165
Parental Involvement	-.09†	0.065	0.165
Parenting Stress (rescore)	-.08†	0.084	0.199
Mental Health - History	.08†	0.1	0.212
Adult Abuse	.08†	0.103	0.212
Childhood Physical Abuse	-.07†	0.173	0.336
PTSD	.06†	0.234	0.428
Mental Health - Anxiety Depression (rescore)	.05†	0.296	0.505
Educational Weaknesses	.05†	0.306	0.505
Adult Sexual Abuse	.05†	0.357	0.561
Attitude Scale	.04†	0.377	0.565
Mental Health - Anxiety Depression	.04†	0.423	0.607
Family Conflict	-.03†	0.529	0.685
Abuse	.03†	0.539	0.685
Childhood Abuse	-.03†	0.539	0.685
Family Support	.02†	0.614	0.75
Childhood Sexual Abuse	.02†	0.682	0.804
Relationship Scale	.02†	0.727	0.827
Educational Strengths	.01†	0.764	0.84
Mental Health - Psychosis	.01†	0.842	0.889
Gambling	-.01†	0.862	0.889
Family Support (rescore)	.01†	0.917	0.917

Reliability Analysis of the WRNA

The reliability analysis reveals substantial variation in internal consistency across the 24 subscales, demonstrating the psychometric strengths and potential areas for improvement within the questionnaire.

Reliability Strength Categories

Excellent Reliability (α/ω > 0.90) Six subscales demonstrated excellent internal consistency: Parental Stress (ω=0.97), Gambling (α=0.97), Gambling Mini-Scale (ω=0.95), Parental Involvement (α=0.94), Self-Efficacy (ω=0.93), and Substance Abuse History (α=0.91). These scales provide highly precise measurement with minimal error variance.

Good Reliability (α/ω = 0.80-0.89) Six subscales showed good reliability: Criminal History Scale (ω=0.85), PTSD (α=0.85), Antisocial Friends (α=0.85), Depression & Anxiety Current (α=0.84), Relationship Scale (ω=0.82), and Anger/Hostility (α=0.81). These scales offer dependable measurement for research purposes.

Acceptable Reliability (α/ω = 0.70-0.79) Six subscales exhibited acceptable reliability: Family Support (α=0.78), Attitudes (α=0.78), Relationship Stability (α=0.77), Mental Health History (α=0.74), Substance Abuse Recent (α=0.71), and Educational Strengths (α=0.70). While adequate for most research applications, these scales contain more measurement error.

Questionable to Poor Reliability (α/ω < 0.70) Five subscales demonstrated suboptimal reliability: Employment Financial (ω=0.66), Housing Safety (α=0.64), Family Conflict (α=0.55), Psychotic Symptoms (α=0.52), and Educational Needs (α=0.51). These scales would benefit from revision to improve measurement precision. Of these, only the Employment Financial scale and the Housing Safety scales are used in the calculation of the risk-score.

Patterns and Implications

The analysis reveals that scales with more items tend to show higher reliability, consistent with psychometric theory. Both gambling-related scales (Gambling and Gambling Mini-Scale) achieved excellent reliability despite the mini version having fewer items, suggesting particularly strong item cohesion within these measures. There was however, a very small proportion of women who declared gambling issues (xx%), which may mean this finding is unstable.

The findings indicate that whilst most of the questionnaire provides reliable measurement, revisions to the lower-performing subscales may be warranted. For research applications, the scales with reliability coefficients above 0.70 can be considered sufficiently reliable, whilst clinical applications might preferably rely on scales with coefficients above 0.80. The inclusion of the Complex-PTSD (ITQ) scale, which demonstrates good reliability (α=0.85), strengthens the questionnaire's utility for trauma-informed assessment.

Table 2: Subscale Reliability measurements

	Scale	Reliability Type	Value
*	Parental Stress	omega	0.97
	Gambling (PGSI)	alpha	0.97
	Gambling (PGSI-mini)	omega	0.95
	Parental Involvement	alpha	0.94
*	Self-Efficacy	omega	0.93
*	Substance Abuse History	alpha	0.91
*	Criminal History Scale	omega	0.85
*	PTSD	alpha	0.85
*	Antisocial Friends	alpha	0.85
*	Depression Anxiety Current	alpha	0.84
	Relationship Scale	omega	0.82
*	Anger Hostility	alpha	0.81
*	Family Support	alpha	0.78
	Attitudes	alpha	0.78
	Relationship Stability	alpha	0.77
	Mental Health History	alpha	0.74
*	Substance Abuse Recent	alpha	0.71
*	Educational Strengths	alpha	0.70
*	Employment Financial	omega	0.66
*	Housing Safety	alpha	0.64
	Family Conflict	alpha	0.55
	Psychotic Symptoms	alpha	0.52
	Educational Needs	alpha	0.51

Items used to calculate the risk-scores are indicated *

Results Key Points

- **Women in higher risk groups were significantly more likely to reoffend** (Low-Risk: 5.2%, High-Risk: 31.8% - a 6.1-fold difference)
- **The WRNA tool effectively predicts both the risk of reoffending AND offending frequency** - higher scores predict both whether someone will reoffend and how many times they will reoffend
- **Nine factors were most strongly linked to reoffending, including:**
  - Criminal history (strongest predictor)
  - Substance abuse (recent, total, and historical)
  - Having friends involved in crime (antisocial associates)
  - Employment and financial problems
  - Anger/hostility issues
- **Higher self-efficacy was a protective factor** - women with greater belief in their ability to succeed had fewer convictions
- **Most sections of the WRNA questionnaire demonstrated good to excellent reliability**
- **As a holistic, trauma-informed tool, the WRNA successfully captures both criminogenic risks and gender-responsive factors** relevant to women's pathways to offending
- **The tool demonstrates robust predictive validity** across multiple dimensions, making it suitable for both risk classification and intervention planning in UK community-based settings

Discussion

This study provides the first comprehensive validation of the Women's Risk Needs Assessment (WRNA) within the UK context, examining both its psychometric properties and predictive validity among justice-involved women accessing community-based services. The findings reveal several important insights regarding the instrument's utility and performance within this specific population and setting.

The predictive validity analyses demonstrate a clear, stepwise relationship between WRNA risk classifications and subsequent offending outcomes. Women classified as High-Risk exhibited reoffending rates approximately 6.1 times higher than those classified as Low-Risk (31.8% versus 5.2%), with moderate and medium categories showing proportionate increases in reoffending rates. This pattern was further corroborated by a significant positive correlation between overall WRNA scores and conviction counts ($\rho = .22$, $p < .001$). The zero-inflated negative binomial regression model revealed that higher WRNA scores were significantly associated with both a reduced probability of zero offences and increased offending frequency among those who do reoffend, demonstrating the instrument's robust dual predictive capacity.

The subscale analysis revealed meaningful correlations between specific WRNA domains and offending outcomes, with nine subscales demonstrating significant associations with conviction counts and eight with binary offending outcomes after adjustment for multiple comparisons. Notably, the analysis confirmed the predictive importance of the "Big Four" criminogenic needs identified in the Risk-Need-Responsivity literature: criminal history, antisocial associates, substance abuse, and employment/financial instability. Criminal History emerged as the strongest predictor ($r = .22/.21$), followed by consistent associations across all substance abuse measures ($r = .18$-$.21$) and antisocial friends ($r = .19$). Additionally, Self-Efficacy demonstrated significant protective effects ($r = -.15/-.14$), highlighting the importance of strengths-based factors in women's desistance processes.

The psychometric evaluation of the WRNA subscales demonstrated considerable variation in internal consistency, with reliability coefficients ranging from excellent to poor. Six subscales achieved excellent reliability ($\alpha/\omega > 0.90$), including Parental Stress, Gambling measures, Parental Involvement, Self-Efficacy, and Substance Abuse History. Twelve subscales demonstrated good to acceptable reliability ($\alpha/\omega = 0.70$-0.89), whilst five subscales exhibited questionable to poor reliability ($\alpha/\omega < 0.70$), notably Employment/Financial, Housing Safety, Family Conflict, Psychotic Symptoms, and Educational Needs. Importantly, only two of the

lower-performing scales (Employment/Financial and Housing Safety) contribute to the overall risk score calculation.

Previous cognitive interviews undertaken for the WRNA (Holdsworth et al., 2023) identified comprehension issues in 17 questions across multiple domains, particularly those containing culturally specific American terminology that required adaptation for UK contexts. These discrepancies likely contributed to increased measurement variance in certain subscales. However, taken together with the current validation findings, the WRNA Version 7 demonstrates promising application within UK settings, whilst acknowledging areas requiring further refinement.

The observed predictive validity of the WRNA aligns with previous research conducted in North American settings (Smith et al., 2009; Van Voorhis et al., 2010), suggesting that the instrument's predictive properties have successfully transferred to the British context despite socio-cultural and criminal justice system differences. This successful cross-cultural application supports Singh et al.'s (2011) assertion that risk assessment instruments can maintain their validity across different jurisdictions when appropriately adapted and implemented.

The stepwise progression in reoffending rates across risk categories reinforces the fundamental premise of the risk principle in offender management (Andrews & Bonta, 2010), which posits that higher-risk individuals require more intensive interventions. The significant relationships observed in both components of the ZINB model demonstrate that the WRNA effectively discriminates between those who do and do not reoffend, whilst also predicting offending frequency among those who do *reoffend*. This dual predictive capacity substantially enhances the clinical utility of the instrument beyond simple risk classification.

Our analysis demonstrated the relevance of some of the 'Central 8' criminogenic needs as primary predictors of female offending, however, as our previous work demonstrates these needs are inextricable linked to 'background' needs such as trauma and mental health (Summers et al., 2024). However, the additional identification of Self-Efficacy as a protective factor underscores the importance of incorporating strengths-based elements in gender-responsive assessment, consistent with the pathways perspective and trauma-informed approaches to women's offending (Brennan et al., 2012; Van Voorhis, 2012).

The variability in subscale reliability echoes findings from other gender-responsive assessment validation studies (Davidson & Chesney-Lind, 2009; Van Voorhis, 2012), which

have noted challenges in measuring certain constructs within female offender populations. The lower reliability of housing and employment-related scales may reflect the situational instability and structural barriers frequently experienced by justice-involved women in the UK (Gelsthorpe & Wright, 2015), which can complicate measurement of these domains. Conversely, the excellent reliability of trauma-related, parenting, and self-efficacy scales reinforces the centrality of these factors in women's pathways to offending, as articulated in the pathways analysis (Brennan et al., 2012).

Implications for Practice and Policy

The findings from this validation study have several important implications for practitioners and policymakers working with justice-involved women in the UK context. Firstly, the demonstrated predictive validity of the WRNA provides empirical support for its implementation within women's centres and other community-based services as a means of identifying women at varying levels of reoffending risk and allocating resources accordingly. The clear differentiation between risk categories enables practitioners to target more intensive interventions towards those at highest risk, consistent with evidence-based principles of effective correctional intervention (Andrews & Bonta, 2010).

Secondly, the comprehensive nature of the WRNA, encompassing both criminogenic needs and gender-responsive factors, allows for a more holistic assessment approach that aligns with the trauma-informed, strengths-based ethos of women's centres. The strong reliability of scales measuring parental stress, self-efficacy, and substance abuse history highlights the instrument's capacity to capture domains particularly relevant to women's desistance journeys. The identification of Self-Efficacy as a protective factor provides empirical support for interventions that build women's confidence and agency, whilst the prominence of substance abuse across all measures reinforces the critical importance of addressing addiction in comprehensive treatment planning.

Thirdly, the variable reliability across subscales suggests that practitioners should exercise caution when interpreting scores from certain domains, particularly Employment/Financial, Housing Safety, Family Conflict, and Psychotic Symptoms. These scales may require supplementary assessment methods to ensure accurate identification of needs. Service providers might consider targeted training to enhance inter-rater reliability for these domains or develop complementary assessment tools to address these specific areas.

From a policy perspective, the findings provide empirical support for the continued development and implementation of gender-responsive assessment approaches within the UK criminal justice system. The demonstrated effectiveness of the WRNA aligns with the recommendations of the Corston Report (2007) and the Female Offender Strategy (Ministry of Justice, 2018), both of which emphasise the importance of addressing women's distinct criminogenic and non-criminogenic needs through tailored assessment and intervention approaches. The successful application of the WRNA within women's centres further reinforces the value of community-based, gender-responsive services as alternatives to custody for women.

Limitations and Future Directions

Several limitations should be considered when interpreting the findings of this study. Firstly, the sample, whilst substantial, was drawn from women accessing services at three specific organisations across seven sites, potentially limiting generalisability to the broader population of justice-involved women in the UK. Future research should aim to validate the WRNA with more diverse samples, including women in custodial settings and those from different geographical regions and demographic backgrounds.

Secondly, the follow-up period of 12 months, whilst practical for initial validation purposes, may not capture longer-term reoffending patterns. Extended follow-up studies would provide valuable insights into the WRNA's predictive validity over more extended time periods, particularly given the typically low base rates of female offending.

The variable reliability across subscales indicates areas for potential refinement of the instrument. Future research should explore item-level analyses to identify problematic items within lower-performing scales and develop revised versions with improved psychometric properties. Additionally, whilst the ZINB model demonstrates significant relationships in both components, further exploration of potential mediating and moderating variables relevant to female offending pathways could enhance understanding of these relationships.

Furthermore, whilst this study establishes the WRNA's predictive validity regarding reoffending frequency and likelihood, it does not explore the severity or nature of offences committed. Future research should examine whether the WRNA accurately predicts offence severity, types of crimes, or progression in criminal behaviour. Additionally, the exclusive reliance on PNC data may have missed certain types of offending (e.g., regulatory breaches), suggesting the need for more comprehensive outcome measures in future validation studies.

Conclusion

This validation study provides robust empirical evidence that the WRNA demonstrates meaningful predictive validity within the UK context, successfully discriminating between women at varying levels of reoffending risk across community-based settings. The instrument's capacity to generate stepwise increases in reoffending rates across risk categories, coupled with significant correlations between assessment scores and subsequent criminal justice outcomes, confirms its utility as a gender-responsive assessment tool for justice-involved women. The dual predictive capacity demonstrated through the zero-inflated negative binomial model—predicting both whether women reoffend and their offending frequency—substantially enhances the clinical utility of the instrument.

The confirmation of established criminogenic risk factors alongside the identification of Self-Efficacy as a protective factor validates both traditional correctional research and gender-responsive approaches to assessment. However, the variable reliability observed across individual subscales, ranging from excellent to questionable psychometric performance, indicates that targeted refinements are necessary to optimise the instrument's measurement precision and clinical utility.

These findings establish a foundation for the continued development and implementation of gender-responsive assessment practices within the British criminal justice system, whilst highlighting specific domains requiring further investigation and potential revision. The validation provides empirical support for the WRNA Version 8 development process and offers practical guidance for practitioners implementing gender-responsive assessment approaches in community-based settings.

References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). LexisNexis/Anderson Publishing.

Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of women offenders: An integrative perspective*. John Wiley & Sons.

Bloom, B., Owen, B., & Covington, S. (2003). *Gender-responsive strategies: Research, practice, and guiding principles for women offenders*. National Institute of Corrections.

Brennan, T., Breitenbach, M., Dieterich, W., Salisbury, E. J., & Van Voorhis, P. (2012). Women's pathways to serious and habitual crime: A person-centered analysis incorporating gender responsive factors. *Criminal Justice and Behaviour*, *39*(11), 1481-1508. https://doi.org/10.1177/0093854812456777

Corston, J. (2007). *The Corston report: A review of women with particular vulnerabilities in the CJS*. Home Office.

Covington, S. S., & Bloom, B. E. (2007). Gender responsive treatment and services in correctional settings. *Women & Therapy*, *29*(3-4), 9-33. https://doi.org/10.1300/J015v29n03_02

Davidson, J. T., & Chesney-Lind, M. (2009). Discounting women: Context matters in risk and need assessment. *Critical Criminology*, *17*(4), 221-245. https://doi.org/10.1007/s10612-009-9084-x

Gelsthorpe, L., Sharpe, G., & Roberts, J. (2007). *Provision for women offenders in the community*. Fawcett Society.

Gelsthorpe, L., & Wright, S. (2015). The context: Women as lawbreakers. In J. Annison, J. Brayford, & J. Deering (Eds.), *Women and criminal justice: From the Corston Report to transforming rehabilitation* (pp. 39-58). Policy Press.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.

Ministry of Justice. (2018). *Female offender strategy*. HM Government.

Radcliffe, P., & Hunter, G. (2016). 'It was a safe place for me to be': Accounts of attending women's community services and moving beyond the offender identity. *British Journal of Criminology*, *56*(5), 976-994. https://doi.org/10.1093/bjc/azv093

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research* (Version 2.3.3) [R package]. Northwestern University. https://CRAN.R-project.org/package=psych

Salisbury, E. J., & Van Voorhis, P. (2009). Gendered pathways: A quantitative investigation of women probationers' paths to incarceration. *Criminal Justice and Behaviour*, *36*(6), 541-566. https://doi.org/10.1177/0093854809334076

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499-513. https://doi.org/10.1016/j.cpr.2010.11.009

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy*, *8*(1), 183-208. https://doi.org/10.1111/j.1745-9133.2009.00551.x

Summers, R., Pemberton, S., Ruprai, D., McKenzie, K., Holdsworth, E., & Long, J. (2024, June). *Examining the relationship between severity & complexity of needs among women involved in the criminal justice system* [online]. University of Birmingham. https://www.birmingham.ac.uk/schools/social-policy/departments/social-policy-sociology-criminology/research/projects/2023/women-crime-justice

Van Voorhis, P. (2012). On behalf of women offenders: Women's place in the science of evidence-based practice. *Criminology & Public Policy*, *11*(2), 111-145. https://doi.org/10.1111/j.1745-9133.2012.00793.x

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behaviour*, *37*(3), 261-288. https://doi.org/10.1177/0093854809357442

Wright, E. M., Van Voorhis, P., Salisbury, E. J., & Bauman, A. (2012). Gender-responsive lessons learned and policy implications for women in prison: A review. *Criminal Justice and Behaviour*, *39*(12), 1612-1632. https://doi.org/10.1177/0093854812451088