

MENTAL HEALTH & RECIDIVISM AMONG WOMEN

**RECIDIVISM AMONG JUSTICE-INVOLVED WOMEN: A COMBINED VARIABLE-  
AND PERSON-CENTERED ANALYTICAL APPROACH TO UNDERSTANDING THE  
ROLE OF MENTAL HEALTH**

by

Cathrine Pettersen

Dissertation submitted to the Faculty of Graduate Studies and Research in Partial Fulfillment of

the requirements for the degree

Doctor of Philosophy

in

Psychology

Carleton University

Ottawa, Ontario

©2021

Cathrine Pettersen

### **Abstract**

The current two-study program of research had two major aims. Study 1 – a comprehensive meta-analysis – was conducted to quantitatively summarize existing empirical knowledge of the relationships between mental health factors and recidivism among justice-involved adult women. Study 2 - a latent class analysis (LCA) – aimed to contribute to the development of mental health-focused correctional profiles that could advance our understanding of justice-involved women and support progress in correctional policy and practice. The meta-analysis included any studies of adult women that examined the association between any mental health factor and recidivism. A comprehensive literature search revealed very sparse empirical data ( $k = 18$ ). The resulting significant albeit modest aggregate effect sizes revealed elevated recidivism rates among women who suffer depression and PTSD, as well as women with any psychiatric diagnosis and more extensive psychiatric histories. Conversely, anxiety, psychoses, and (unspecified) personality disorders were not significantly related to recidivism. A history of externalizing behaviours such as self-harming and suicide attempts did not predict recidivism. Study 2 - the mental health-focused LCA - involved a sample of 920 women incarcerated in Maine state prisons with complete risk/need assessments including mental health data on the SPIn-W (Orbis Partners, 2006). Three latent classes best described the sample in terms of nature, severity, and complexity of mental health needs; Class 1 “severe needs” (30%), Class 2 “moderate needs with severe externalizing behaviours” (20%), and Class 3 “low needs” (50%). Women in the severe and moderate needs classes had higher recidivism rates than women in the low needs class. The findings support the recommendations of contemporary women-centered correctional scholars; holistic, trauma- and mental health-informed assessment, management, and treatment approaches are required to meet the needs of justice-involved adult women.

## **Acknowledgements**

I would like to thank my supervisor Dr. Shelley Brown for her guidance, unwavering support, and continued encouragement through difficult times. Working with you gave me the knowledge, insights, and most importantly the courage to continue on in the face of overwhelming personal challenges and setbacks. Thanks to your mentorship I was able to successfully navigate through the program and the many complexities of the research process. This dissertation would not have been possible without you.

I would like to thank my committee members Dr. Julie Blais and Dr. Ralph Serin for your insightful feedback and thoughtful suggestions that tremendously impacted the quality of my program of research. I would further like to thank my internal examiner, Dr. Kathryn Kelly, and my external examiner, Dr. Margo Watt, for dedicating time examining my dissertation document. All of my fellow colleagues in the Gender and Crime Lab deserve special mention for sharing their invaluable know-hows and providing support throughout the research process; Julie Goodwin-a special thanks to you for your contribution to project during preparatory phases of the meta-analysis. Megan Wagstaff, Colleen Robb, and Kayla Wanamaker- thank you for offering your time as well as technical and analytical expertise, even during your most busy of times.

A sincere thank you to the Maine department of corrections, Orbis Partners, Inc and Dr. David Robinson in particular, for providing your data, including the Service Planning Instrument for Women (SPIn-W) assessments, providing the training necessary, and for making yourself available for consultation and guidance at critical points during the research process.

Good friends and family members provided me with the motivation and endurance to finish this project. A special thank you to my ex-husband Duane for offering personal support

and for opening your doors to me during challenging, uncertain times. Tammy- you've been my faithful friend for well over a decade; during the most impossible of times you were there for me in the most incredible ways. You'll always be close to my heart. Had it not been for you, I would not have completed this project. Genevieve-always there to cheer me on and help me back on track whenever I waivered – it won't be forgotten. Dr. Henriette Bergström- Henri my friend- thank you for being there with sound advice to help me navigate the world of academia, share your knowledge and expertise on the most complex of issues facing correctional psychology today, and for always answering to my call when I needed it most.

Michael, if it wasn't for you I would not have succeeded in obtaining my psychology doctorate. You too facilitated my hard-headed pursuit of this achievement in the most surprising and inspiring ways. I wish for you only the very best in this life.

A owe a special debt to my now retired long-term psychoanalyst, Dr. C. L., for the unconditional positive regard you offered me throughout my difficult journey from mental illness toward mental health and a better life. Just this once, your prognosis was wrong; I was able to finish my dissertation and remain employed in the face of all obstacles. Thank you for that moment of doubt in my capacities to cope; without that kick right in the behind I would never have obtained the level of commitment necessary to realize my ambitions.

I would like to thank my father for his unshakable belief in my ability to hang on to the rudder during the stormiest of seas and for being my motivation to always press on.

## Table of Contents

Abstract.....	i
Acknowledgements.....	i
List of Tables .....	i
List of Figures.....	i
List of Appendices .....	i
Recidivism Among Justice-Involved Women: Combined Variable- and Person-Centered Analytical Approaches to Understanding the Role of Mental Health .....	1
Chapter 1: General Introduction.....	1
Mental Health as a Specific Responsivity Issue.....	3
The Mental Health Status of Justice-Involved Women .....	4
Prevalence of Mental Disorders among Federally Incarcerated Women.....	6
Theoretical Perspectives: Gender-neutral and Gender-responsive Approaches .....	12
The Gender-Neutral Perspective .....	12
The Gender-Responsive Perspective.....	17
Study Objectives .....	20
Chapter 2.....	22
Mental Health and Adult Women’s Recidivism: A Meta-Analysis .....	22
A Brief Overview of the Literature .....	22
Symptom Clusters, Mental Disorders, and Theories of Risk Relevance .....	22
Responsivity: Revisiting the Risk-relevance of Mental Disorder in Treatment Contexts .....	25
Women’s mental health and recidivism .....	29
Summary .....	35
The Current Meta-Analysis.....	36
Scope of the Meta-Analysis .....	37
A–Priori Exclusion Criteria.....	38
Research Questions .....	39
Method .....	41

Meta-analysis Protocol Registration .....	42
Inclusion Criteria.....	42
Exclusion Criteria.....	43
Additional Post-hoc Exclusion Criteria.....	43
Literature Search .....	44
Search Terms.....	47
Coding.....	49
The Complete Coding Protocol: Overall Structure and Main Features .....	51
Inter-Rater Reliability .....	53
Indexing the Average Effect: Choice of Effect Size Measures.....	56
Study Heterogeneity and Outliers .....	62
Moderators .....	64
Independence of Observations .....	66
Preliminary Findings: Literature Search and Inter-rater Reliability.....	67
Literature Search Summary.....	67
Summary Report .....	70
Results of Inter-rater Reliability Analyses .....	72
Main Results: What the Quantitative Synthesis Tells Us .....	79
Depression and General Recidivism .....	80
Anxiety and General Recidivism .....	85
Post-traumatic stress disorder (PTSD) and General Recidivism .....	87
Psychosis and General Recidivism .....	88
Personality disorder (PD) and General Recidivism .....	89
Any Mental Disorder and General Recidivism.....	91
Self-harm and Suicidality and General Recidivism .....	94
Psychiatric History and General Recidivism .....	95
Discussion.....	97
Limitations .....	105
Future Research.....	108
A Note on Practice and Policy .....	111

Summary .....	112
Chapter 3 .....	114
Study 2: Mental Health Profiles of Justice-involved Adult Women: A LCA of the SPIn-W ....	114
Fundamentals of LCA and LPA: A brief Overview .....	116
Previous LCA and LPA: Mental Health Profiles of Justice-involved Women.....	118
Trauma and Adversity in Childhood among Justice-Involved Adult Women.....	126
The Current Study .....	128
Research Questions .....	129
Method .....	131
Participants .....	131
Measures.....	135
<i>The Service Planning Instrument for Women (SPIn-W; Orbis Partners, 2006)</i> .....	135
<i>Mental Health Indicators: Items Derived from the Full Assessment</i> .....	137
<i>SPIn-W: Pre-Screen (see Appendix J for copy)</i> .....	139
<i>Adverse Childhood Experiences Scale (ACEs)</i> .....	140
<i>Recidivism</i> .....	142
Procedure & Acquisition of Data .....	143
Analytical Approach: Preparatory & Exploratory Analyses.....	143
LCA Procedures: Preparations for a Multi-Stage LCA Analysis .....	147
Evaluation of Model Fit and Latent Class Solutions .....	150
Results.....	151
Preliminary Analyses: Pre-LCA Explorations of Mental Health & Adversity.....	151
ACEs .....	153
LCA Results .....	155
Model 1: What the 14 Mental Health Indicators Revealed.....	156
Model 2: The Addition of the Pre-Screen Risk Score Covariate .....	160
Model 3: Using Auxiliaries to Better Understand Latent Classes .....	162
Model 4: The Complex Final Model.....	164
Post-Hoc Exploratory Analyses .....	168
Linking Mental Health and Recidivism: Additional Class Comparisons .....	171

Comprehensive Class Descriptions: Who Are the Women from Maine .....	174
Class 1: Severe needs .....	174
Class 2: Moderate needs with severe externalizing behaviours.....	175
Class 3: Low needs.....	175
Conclusion.....	176
Discussion.....	177
Future Research.....	184
Implications for Practice .....	187
Chapter 4.....	190
General Discussion .....	190
Summary of Main Findings .....	191
Limitations of the Current Program of Study.....	196
Implications for Future Research .....	199
Implications for Policy and Practice .....	203
Concluding Remarks .....	207
References.....	209

## List of Tables

<b>Table 1.</b> <i>Prevalence of Mental Disorders among Federally Sentenced Women and Men in Canada (2015-2019)</i> .....	11
<b>Table 2.</b> <i>Mental Health and Recidivism among Adult Justice-Involved Women: Overview of Selected Studies and Findings</i> .....	31
<b>Table 3.</b> <i>Inter-Rater Agreement (Reliability): Study &amp; Sample Characteristics (Kappa &amp; ICC)</i> .....	78
<b>Table 4.</b> <i>Inter-Rater Agreement (Reliability): Effect Size Data (Kappa &amp; ICC)</i> .....	79
<b>Table 5.</b> <i>Depression: Study and Sample Characteristics (k = 11)</i> .....	85
<b>Table 6.</b> <i>Depression: Recidivism, Attrition, Sample Sizes, &amp; Cohen's d Effect Sizes (95% CI)</i> .....	86
<b>Table 7.</b> <i>All Mental Health Predictors &amp; General Recidivism: Aggregate Effect Sizes &amp; Between Study Variability</i> .....	82
<b>Table 8.</b> <i>All Supplementary Categorical Moderator Analyses: Aggregated Fixed Cohen's d Effects: All Predictors &amp; General Recidivism</i> .....	83
<b>Table 9.</b> <i>Anxiety (k = 11) and PTSD (k = 4): Study and Sample Characteristics</i> .....	85
<b>Table 10.</b> <i>Anxiety, PTSD, and General Recidivism: Individual Cohen's d Effect Sizes &amp; Related Outcome and Sample Size Data (k = 11)</i> .....	86
<b>Table 11.</b> <i>Psychosis: Study and Sample Characteristics (k = 7)</i> .....	88
<b>Table 12.</b> <i>Psychosis &amp; General Recidivism: Individual Cohen's d Effect Sizes &amp; Related Outcome &amp; Sample Size Data (k = 7)</i> .....	89
<b>Table 13.</b> <i>PD &amp; General Study and Sample Characteristics (k = 4)</i> .....	90
<b>Table 14.</b> <i>PD &amp; General Recidivism: Cohen's d Effect Sizes, Outcomes, &amp; Sample Sizes (k = 4)</i> .....	90
<b>Table 15.</b> <i>Any Mental Disorder (MD): Basic Study &amp; Sample Characteristics</i> .....	91
<b>Table 16.</b> <i>Any Mental Disorder &amp; General Recidivism: Individual Cohen's d Effect Sizes &amp; Related Outcome &amp; Sample Size Data (k = 6)</i> .....	92
<b>Table 17.</b> <i>Self-harm and Suicidality: Basic Study &amp; Sample Characteristics (k = 4)</i> .....	93

<b>Table 18.</b> <i>Self-harm, Suicidality, &amp; General Recidivism: Individual Cohen’s d Effect Sizes, Outcomes, &amp; Sample Sizes (k = 4)</i> .....	93
<b>Table 19.</b> <i>Psychiatric History: Basic Study &amp; Sample Characteristics (k = 4)</i> .....	94
<b>Table 20.</b> <i>Psychiatric History &amp; General Recidivism: Individual Cohen’s d Effect Sizes, Outcomes &amp; Sample Sizes (k = 3)</i> .....	95
<b>Table 21.</b> <i>Latent Class Solutions (Number of Classes) in Previous LCA and LPA</i> .....	118
<b>Table 22.</b> <i>Sample Characteristics: Means (SD), Frequencies, and (Sub)sample Sizes</i> .....	132
<b>Table 23.</b> <i>LCA Indicators &amp; Additional Mental Health Variables: Sample Proportions</i> .....	144
<b>Table 24.</b> <i>Adverse Childhood Experiences (ACEs): Means (SD) &amp; Endorsement Frequencies</i> .....	153
<b>Table 25.</b> <i>Fit Indices, Entropy, &amp; LRT: All Models with One- to Four-Class Solutions</i> .....	157
<b>Table 26.</b> <i>Class Proportions (n/%) for All Models with Two-, Three-, and Four-Class Solutions</i> .....	159
<b>Table 27.</b> <i>Auxiliary Variables ACEs, Age, &amp; Race in Models 3 and 4: M (SE) and Chi-Square Difference Tests (n = 625)</i> .....	164
<b>Table 28.</b> <i>Model 4 with Three Classes: Mental Health Indicator Probabilities (n = 625)</i> .....	165
<b>Table 29.</b> <i>Pairwise Comparisons of Latent Classes: MH Indicators and Co-Occurring Disorders</i> .....	171
<b>Table 30.</b> <i>Comparing Latent Classes on Returns to Custody across Four Follow-up Times</i> .....	172

## List of Figures

<b>Figure 1.</b> <i>Reliability- Manual, Coder Training, &amp; Inter-rater Reliability Analyses</i> .....	54
<b>Figure 2.</b> <i>Literature Review Flow Chart: Academic Databases</i> .....	73
<b>Figure 3.</b> <i>Literature Review Flow Chart: Alternative Sources</i> .....	74
<b>Figure 4.</b> <i>Literature Review Flow Chart: All Included Studies from all Sources</i> .....	75
<b>Figure 5.</b> <i>The SPIn-W Wheel (reconstruction; Jones &amp; Robinson, 2018)</i> .....	137
<b>Figure 6.</b> <i>Model 4: Mental Health Indicator Probabilities (Percentages): The Final Three- Class Solution</i> .....	166

## List of Appendices

<b>Appendix A.</b> Non-English Sources and Search Terms.....	244
<b>Appendix B.</b> Meta-analysis Coding Protocol.....	247
<b>Appendix C.</b> Risk of Bias in Observational Research.....	261
<b>Appendix D.</b> Coding, Study, and Data Selection: Post-Eligibility Screening.....	265
<b>Appendix E.</b> Meta-analytic Formulae.....	275
<b>Appendix F.</b> Individual Effect Size Selections: Documenting Judgment Calls.....	278
<b>Appendix G.</b> List of Authors Contacted and Inquiry Results.....	289
<b>Appendix H.</b> References for Rejected Studies and Reasons for Rejection.....	291
<b>Appendix I.</b> Inter-rated Reliability Analyses: Exclusions and Strategic Approaches.....	321
<b>Appendix J.</b> SPIn-W Pre-Screen Assessment (Orbis Partners, 2006).....	322
<b>Appendix K.</b> Adverse Childhood Experiences Scale (ACES): 12-Item SPIn-W Proxy Scale.....	328
<b>Appendix L.</b> Additional LCA Models Tested.....	329
<b>Appendix M.</b> Memorandum of Understanding (MOU).....	332
<b>Appendix N.</b> Certification of Institutional Ethics Clearance.....	333
<b>Appendix O.</b> SPIn-W Mental Health Items: Indicator, Scaling, Risk Scores, & LCA Utility.....	335
<b>Appendix P.</b> Mplus Syntax: The Final Model 4: Three-Class Solution.....	336

## **Recidivism Among Justice-Involved Women: Combined Variable- and Person-Centered Analytical Approaches to Understanding the Role of Mental Health**

### **Chapter 1: General Introduction**

The number of women serving federal sentences in Canada has increased dramatically in the past decade (Public Safety, 2019). Indeed, the total number of women admitted to federal correctional institutions has increased by 26% from 266 in 2012/13 to 357 in 2017/18 (Public Safety, 2019). Although only a minority of all violent offenses are committed by women, 53% ( $n = 754$ ) of women serving federal community and custodial sentences during the fiscal year of 2018/19 were sentenced for a violent offense (Public Safety, 2019). Recent research suggests that a large proportion of incarcerated women in Canada suffer from mental health problems at intake into correctional institutions and continue to experience such problems while serving their sentences (Derkzen et al., 2012; OCI, 2019). In fact, the most recent estimates from the Office of the Correctional Investigator's (OCI) Annual Report indicate that 79% of federally incarcerated women in Canada suffer at least one mental disorder and that one third of the population meet criteria for post-traumatic stress disorder (OCI, 2019). The study of the mental health and trauma-related challenges faced by justice-involved women is of paramount importance to maintain progress in future research and thereby increase our scientific knowledge in this area, and relatedly, future correctional practice and policy.

First, mental health problems are one of the greatest obstacles to effective and ethical management of incarcerated populations faced by correctional systems today (Derkzen et al., 2012). Mental health impacts the correctional agencies' and institutions' ability to reach their primary goals in several ways. First, the goal of effective, safe, and humane management can be impaired by high rates of severe mental health problems by increasing both self- and other-

directed aggression, violence, as well as institutional misconducts in general; Rates of self-harming behaviours, including suicide attempts are extremely high in custody settings.

Correctional agencies such as the Correctional Service of Canada (CSC) oftentimes do not have adequate facilities, staff, or the expertise to appropriately manage individuals who chronically self-injure (OCI, 2019), resulting in unacceptable, punitive correctional responses (e.g., the use of static security measures such as solitary confinement – also known as segregation – in an attempt to regain control of a disruptive mentally ill client; see Saper's 2008 report on a young, severely mentally disturbed woman -Ashley Smith- who died as a result of prolonged segregation). The additional stress, deprivation, and re-traumatization experiences of all parties fuel a combustible, unsafe environment contrary to rehabilitative and humanitarian goals. Mentally disordered women may also be violent against others, which obviously leads to further management problems.

Second, a main goal of the criminal justice system itself, rehabilitation, can be jeopardized by severe mental illness. Mental disorder may constitute a specific responsibility issue, meaning that offenders with untreated mental health problems may not be able to adequately engage with, or respond to, correctional programs aimed at reducing recidivism (McCormick et al., 2015). Consider the concept of specific responsibility in the context of a weight-loss program intended for morbidly obese individuals needing to lose weight to avoid negative health outcomes. The obvious solution seems to be proper diet and exercise. However, if your participants broke both their legs leading up to treatment commencement, this medical emergency would need to be addressed before the weight loss program could be effective. Mental health obstacles may operate similarly to a broken leg; without first addressing them, programming will remain ineffective. If this were the case, mental health problems would

constitute an obstacle to rehabilitation and, ultimately public safety. Specific responsivity in this context is discussed at length in the below section.

### **Mental Health as a Specific Responsivity Issue**

Lastly, a direct or indirect link between mental illness and recidivism remains a possibility. The association between mental health and recidivism has only been examined in a limited number of studies of justice-involved adult women (King et al., 2018; Van Voorhis et al., 2010). The results of these women-specific studies are equivocal, with some findings suggesting that severe mental health problems (e.g., mood disorders) are associated with increases in recidivism (Van Voorhis et al., 2010), while others have uncovered no relationship between mental health and reoffending (Hubbard & Pratt, 2002). Further, some researchers have also documented the opposite; an inverse association between mental health and recidivism (Blanchette, 1996; Blanchette & Motiuk, 1996). Interestingly, and in accordance with traditional and contemporary women-centered correctional theory, findings suggesting that poor mental health leads to poor behavioural outcomes including recidivism are juxtaposed to mainstream findings from studies of justice-involved men (Bonta et al., 1998; Bonta et al., 2014). In research focused exclusively on men, virtually no adequate empirical support has been uncovered supporting the contention that poor mental health is related to increased risk and recidivism.

In this chapter, the following literature will be reviewed. First, the prevalence and nature of the mental health challenges faced by justice-involved adult women in North American jurisdictions are reviewed; where appropriate, brief comparisons to studies of justice-involved men will be made. Second, a brief overview of traditional correctional research perspectives on gender and recidivism risk- often referred to as the gender-neutral approach- will be provided. Gender-neutral refers to the assumption that risk and need factors are the same for women and

men both in nature and saliency, as well as in terms of viable treatment approaches. As it pertains to recidivism risk, gender-neutral researchers do not recognize mental health as a criminogenic need (Andrews & Bonta, 2017). Third, a brief discussion of mental health problems as a potential specific responsivity issue will follow. In the current context, specific responsivity refers to gender-neutral theorists and researchers' suggestion that mental health may affect individual clients' response to correctional treatment. Third, the alternative gender-responsive perspective and its approach to the study of mental health in justice-involved populations in general and women in particular, is introduced. The gender-responsive perspective proposes precisely the opposite theoretical stance on risk factors and gender: what constitutes a risk factor, the magnitude of its effects on recidivism, and the best way to reduce recidivism risk through intervention, may differ depending on gender. Lastly, I will briefly describe the main, over-arching goals of the current study, followed by a brief guide to the organization of the remainder of this document.

### **The Mental Health Status of Justice-Involved Women**

Well-established gender-responsive theory, consistently report elevated rates of mental health disorders and symptoms among women relative to men, as well as findings that suggest gender-dependent effects of some risk factors (i.e., substance abuse predicts recidivism for both genders, but the magnitude of the effect is larger among women) support gender-responsive scholars' position that the population of justice-involved women may require special consideration.

The rates of mental disorder (e.g., internalizing disorders such as anxiety) as well as non-diagnostic indicators and behavioural expressions of mental distress and dysfunction have been consistently found to be elevated in samples of justice-involved women as compared to justice-

involved men (Blanchette & Motiuk, 1996; Brinded et al., 2001; Coolidge et al., 2011; Derkzen et al., 2014; James & Glaze, 2006; Magaletta et al., 2009). These findings are not limited to North-America; replications have also been reported in Australia (Butler et al., 2005; Butler et al., 2006) and in the United Kingdom (U.K; Brugha et al., 2005). In addition, the rate of exposure to trauma during important developmental phases – from infancy to late adolescence – are very high among women relative to their male counterparts; girls and young women are more frequently physically, sexually, and emotionally victimized (Coolidge et al., 2011; Messina, et al., 2006; Wanamaker et al., 2016; Wasserman et al., 2005).

Retaining primary focus on North-American jurisdictions, James and Glaze (2006) examined the prevalence rates of mental health problems in 705, 600 offenders in state prisons, 78, 800 in federal prisons, and 479, 900 in local jails in the United States. Mental health problems were defined as a) clinical diagnoses of mental disorders using criteria outlined in a recent version of the Diagnostic and Statistical Manual of Mental Disorders (Fourth edition; DSM-IV; American Psychiatric Association [APA], 2000), and/or b) treatment by a mental health professional in the past 12 months. The prevalence of mental disorder diagnoses were higher among incarcerated women than among their male counterparts: 73% vs. 55% in state prisons and 75% vs. 63% in local jails met diagnostic criteria.

Importantly, evidence for this gender gap is not equivocal across all mental disorders. Previous research suggests that the clearest gender differences occur for specific internalizing disorders such as generalized anxiety disorder (GAD) and post-traumatic stress disorder (PTSD; Belknap & Holsinger, 2006; Brinded et al., 2001). Comorbidity of mental disorders and substance use disorders have also been found to be more common among women than men (Blume, 1997). Conversely, Fazel and Seewald (2012)- in a systematic review and meta-

regression analysis of psychotic disorders and major depression involving 28,361 men and 5,227 women incarcerated in 24 different countries reported comparable rates of psychotic disorders and major depression for women and men. The authors did not examine conditions such as mood disorders, anxiety disorders, or personality disorders in their review.

Additionally, some disorders are relatively rare and perhaps more difficult to accurately identify and diagnose. Thus, very little is known about disorders like binge eating disorder, thought and adjustment disorders, and many personality disorders such as schizoid, schizoaffective, and histrionic personality styles.

The somewhat inconsistent findings regarding gender differences in mental health status among justice-involved women and men prevent broad, firm conclusions about gender and mental health in its totality. Nevertheless, sufficient evidence has been generated to support a growing consensus in the scientific literature that women's correctional needs and in particular their mental health needs, differ meaningfully from those of men. The resulting conclusion has been that these gender-salient needs require a research and treatment approach that is tailored to women (i.e., gender-responsive research and treatment approach; Blanchette & Brown, 2006; Derkzen et al., 2014; King et al., 2018; Van Voorhis et al., 2010).

### **Prevalence of Mental Disorders among Federally Incarcerated Women**

As indicated above, mental health problems are the norm rather than the exception among women offenders (Bloom & Covington, 2008) and studies of women report higher rates than studies of men for most disorders (Brown et al., 2018; Beaudette et al., 2015).

To obtain the most reliable estimates, the academic research literature as well as major government reports (e.g., OCI's Annual Reports) were searched and the three studies with the most representative, largest sample sizes that were limited to federally incarcerated women and

men, and that were recently published were selected Brown et al. (2018; women), OCI (2019; women), and Beaudette et al. (2015; men) were selected. All studies based their rates of diagnoses included in and in accordance with the Diagnostic and Statistical Manual of Mental Disorders (4<sup>th</sup> revised version; DSM-IV-TR; American Psychiatric Association [APA], 2000). All samples were Canadian. Participants were federally incarcerated men and women, with the exception of 4.2% of the overall total sample for all three studies ( $N = 2,032$ ), all of whom were women who were primarily supervised in the community. Importantly, previous research has found that the prevalence rates of mental disorders among justice-involved populations in Canada and the United States are largely comparable (e.g., Corrado et al., 2000).

### **Studies of Women**

Brown and colleagues' (2018) study utilized a sample of 246 women federally sentenced in Canada between 2016 and 2017. Sixty-five percent of these women were from a federal women's prison and represented 23% of the total federally incarcerated women population in Canada; 86 were from an federal intake sample of women incarcerated for no more than 90 days. Structured clinical interviews tailored to the DSM-IV-TR (APA, 2000) were used to make diagnoses. The OCI's (2019) Annual Report covers the fiscal year of 2017/18. The prevalence estimates reported here were based on the entire federally incarcerated population in Canada during that period ( $N = 676$ ) and as such likely represents the most accurate estimates available. Nevertheless it is worthwhile to note that the soundness of assessments and the disorder selection criteria were not reported.

### **Studies of Men**

Beaudette and colleagues (2015) conducted a national prevalence study of mental disorders among federally sentenced men at the intake into federal prisons. The volunteer

participants differed from the overall sample initially invited to the study; they had lower risk, less intense needs, and lower security classifications. In line with Brown et al.'s (2018) approach, Beaudette and colleagues also relied on structured clinical interviews tailored to the DSM-IV-TR (APA, 2000).

The studies reported on a vast range of mental disorders. Table 1 below presents prevalence rates for mental disorders that were selected here because they a) affected a large proportion of the sample(s), b) were associated with severe dysfunction, c) were theoretically or empirically of special relevance to correctional contexts, d) represented unique features of women's mental health profiles (e.g., eating disorders), or e) had high negative impact potential on the individual's success in institutional settings and reaching correctional goals. In the current section, only disorders with extremely high prevalence rates among women, as well as disorders characterized by very large gender differences will be discussed. See Table 1 for additional details.

First, it is worth directing attention to the fact that for all disorders, women's rates exceeded those of men. However, not all differences were large in magnitude, and not all could reliably attributed solely to innate gender differences. In terms of broad measurement of mental dysfunction, it was found that 79% of women currently met diagnostic criteria for at least one mental disorder (any). Men's rates were marginally lower (6.4%). In terms of lifetime diagnoses of any mood disorder, more than 50% of women met criteria; men's rates were approximately 20% lower. Notably however, the gender gap all but closed when focus was directed to current disorders; women and men differed by about 5%. In line with previous research, the most common disorder type among women was anxiety disorders, with just over half of all women meeting both criteria both in terms of lifetime and current disorders. Again, men's rates were

quite noticeably lower, affecting about 33% of the sample. As anticipated, PTSD was a very common anxiety disorder among women, with just under 30% affected by this set of symptoms. Men's rate of PTSD was 50% lower than that of women.<sup>1</sup> Eating disorders were as much as ten times higher among women than men, though less is known about these disorders impact on and relevance to adequate functioning in correctional settings. As previous studies have repeatedly indicated, borderline personality disorder (BPD) is a personality disorder diagnosis overwhelmingly applied to justice-involved women; Thirty-three percent of women had a lifetime diagnosis, while the rate for men was reduced by just over 50%. Finally, also in accordance with previously reported findings, the rates of comorbidity between any mental disorder and a substance abuse/dependence disorder differed wildly between genders; roughly 80% of women versus just over 50% of men were affected by the difficult combination of these disorder types. Rates of comorbid mental disorder and one of two personality disorders (APD or BPD) are reported in Table 1, but are not subject to direct gender comparisons, as inclusion criteria differed between men and women. Nonetheless, it is concerning to note that over the life span, about 63% of women have a mental disorder as well as either APD, BPD, or both.

There are noteworthy limitations to both the prevalence rates and gender comparisons conducted above. First, all initially reviewed studies, including the three selected here, failed to report on other arguably highly relevant disorders such as ADHD as well as well-established non-diagnostic indicators of mental health problems such as self-harm and suicidal ideation or attempts. These omissions necessarily result in underestimated overall prevalence rates, and even more pressing, incomplete mental health profiles that would seriously suppress any correctional

---

<sup>1</sup> PTSD was an anxiety disorder in the DMS-IV-TR (APA, 2000). In the current DSM-V, it is no longer considered an anxiety disorder but rather a trauma and stressor related disorder.

agency's ability to adequately support women suffering these conditions. In addition, correctional scholars often fail to consider alternative explanations for apparent gender differences, and the current context is no exception. For example, we know that proportionately fewer mental health resources are directed toward federally incarcerated men than women, resulting in fewer assessment opportunities to detect mental disorder among men. Further, given what is generally known about gender differences in terms of openness about personal-emotional challenges, men may be more likely to hide or understate mental health symptomology on account of a preference for traditional gender roles, very likely enhanced in an environment favoring traditional or even hostile masculinity. Finally, it seems plausible that given that on average, men may express emotions and distress in less overt ways than women. Thus, their mental health challenges may be more difficult to discover (Belknap & Holsinger, 2006; Muller & Kempes, 2016).

Despite these limitations, a few conclusions can confidently be drawn. First, the vast majority of federally sentenced and incarcerated women suffer a diagnosed mental illness; the need to address these serious health concerns in correctional institutions as well as in community settings are undeniable. The most common current type was anxiety disorders, in particular PTSD. BPD is also prominent in women's mental health profiles, and this is noteworthy given its permanency, the degree of persistent dysfunction with which it is associated, its association with increased risk and recidivism, and the challenge it presents in terms of effective treatment (Lawson et al., 2010; Newhill et al., 2009; Ross & Babcock, 2009). Finally, though variance along gender lines were not always pronounced, differences large enough in magnitude to reasonably be attributable to natural variations in the patterned mental health profiles of women and men were present but limited to a few specific disorders.

**Table 1.**

*Prevalence of Mental Disorders among Federally Sentenced Women and Men in Canada (2015-2019)*

Mental disorder (MD)	Women ( <i>n</i> = 246 <sup>a</sup> ; <i>N</i> = 676 <sup>c</sup> )		Men ( <i>N</i> = 1,110) <sup>b</sup>	
	Lifetime*	Current	Lifetime*	Current
Major Mental disorder (MMD) <sup>d</sup>	52.0%	17.9%	-	12.4%
Any Mental Disorder (Any MD)	88.2%	79.2% <sup>e</sup>	81.0%	72.8%
Mood disorders	<b>53.3%</b>	<b>22.1%</b> <sup>e</sup>	<b>30.2%</b>	<b>16.9%</b>
Major depressive disorder	36.6%	9.8%	18.0%	7.4%
Bipolar disorders (All)	10.6%	4.1%	6.1%	3.9%
Anxiety disorders	<b>52.8%</b>	<b>54.2%</b> <sup>e</sup>	<b>34.1%</b>	<b>29.5%</b>
Post-traumatic stress disorder (PTSD)	<b>28.9%</b>	<b>27.2%</b>	<b>13.4%</b>	<b>11.0%</b>
Psychotic disorders (All)	8.1%	4.6% <sup>e</sup>	4.7%	3.3%
Eating Disorders (All)	13.8%	11.0% <sup>e</sup>	1.4%	0.8%
Borderline personality disorder (BPD)	<b>33.3%</b> <sup>e</sup>	N/A <sup>e</sup>	<b>15.9%</b>	N/A <sup>e</sup>
Comorbid MD and SA	<b>81.1%</b>	-	<b>52.6%</b> <sup>f</sup>	-
Comorbid MD and APD/BPD	<b>63.1%</b>	-	<b>36.0%</b> <sup>g</sup>	-

*Note.* Disorders included in this table were selected to include only the most relevant in terms of percentage of population/sample affected, severity (e.g., psychotic disorders), impact on correctional environment (e.g., BPD), and especially relevant to women’s mental health profiles (e.g., PTSD). MD = mental disorder; MMD = major mental disorder; PTSD = post-traumatic stress disorder. BPD = borderline personality disorder; APD = antisocial personality disorder; SA = substance/alcohol Abuse/dependence disorders. <sup>a</sup>Brown et al. (2018). <sup>b</sup>Beaudette et al. (2015). <sup>c</sup>OCI (2019); when multiple sources reported prevalence rates for the same disorder, this report was preferred as rates were based on entire population of federally incarcerated women rather than a subsample. <sup>d</sup>Major mental disorder included major depressive disorder, bipolar disorders I and II, and any psychotic disorder. <sup>e</sup>Current rates are not relevant to BPD diagnoses, as personality disorders are considered permanent (they are managed, not cured). <sup>f</sup>This proportion excludes comorbid APD. <sup>g</sup>Excluded BPD but included SA. “-” = not available./reported. Noteworthy gender differences are **bolded**. \*Lifetime prevalence estimates are approximate.

### **Theoretical Perspectives: Gender-neutral and Gender-responsive Approaches**

As mentioned above, two broad theories of criminal behaviour are currently dominating the research literature on women offenders: the gender-neutral perspective and the gender-responsive perspective. This section will briefly compare these perspectives to situate the current program of study within the broader context of contemporary research literature on justice-involved women.

#### **The Gender-Neutral Perspective**

The most common gender-neutral paradigm is called the risk-need-responsivity (RNR) model (Bonta & Andrews, 2017). This model is based on the general personality and cognitive social learning (GPCSL) theory of crime (Bonta & Andrews, 2017), a multidisciplinary approach with psychology at the forefront. Briefly, GPCSL theorists posit that an individual will engage in crime if and when the rewards for doing so outweigh the costs (i.e., the decision to engage in crime is based on a rational cost-benefit analysis). Crucially, the RNR perspective was constructed based on research that used samples consisting either entirely or predominantly of justice-involved men. With regards to gender and crime, the general assumption of this correctional model is that pathways into crime and the factors that influence recidivism are gender-neutral. If a factor is an established dynamic risk factor for recidivism and considered a viable treatment target for men (i.e., a criminogenic need) then it is for women as well. If a dynamic psychosocial factor has not been found to be risk-relevant and treatable among men, then it will also be irrelevant to risk and thus not targeted in treatment for women.

The RNR model has become the dominant theory guiding contemporary correctional programming (Bonta & Andrews, 2017; Peterson-Badali et al., 2015). Within this model, also

referred to as the Canadian model (Van Voorhis et al., 2010), three principles have been emphasized: risk, need, and responsivity.

First, the risk principle states that future offending can be predicted from the offender's level of risk and involves assigning the highest intensity interventions to the highest risk cases (i.e., matching the level of treatment to the level of risk). Second, the need principle states that dynamic criminogenic needs—those (gender-neutral) risk factors that have been found, through empirical quantitative research, to directly predict recidivism—should be the factors targeted in correctional treatment. The gender-neutral risk factors (i.e., criminogenic needs) are most commonly summarized in terms of three distinct domains; antisocial behavioural history (i.e., criminal history), antisocial personality pattern, antisocial attitudes (i.e., attitudes supportive or criminal behaviour), antisocial peers (i.e., criminal friends), substance abuse, and level of rewards and satisfaction in the areas of family and marital relationships, school and work, and leisure and recreation (Andrews et al., 2011). Notably, mental health is not considered an important risk/need factor among gender-neutral scholars, as it is often left out of gender-neutral assessments entirely (e.g., Level of Service/Case Management Inventory [LS/CMI]; Andrews et al., 2004), or is assessed in a very limited way (LSI-R; Andrews & Bonta, 1995; Wright et al., 2007). Lastly, the third principle of responsivity involves applying treatment in a way that will be most effective at reducing recidivism. Specifically, when treatment programs and modalities are designed, they must account for individual idiosyncrasies that are likely to affect the individual client's ability to participate fully in, and benefit from, the intervention. Such individual factors can include learning styles, learning disorders, first language, personality disorders, physical health problems, and mental disorders.

It is notable that within the RNR literature, the responsivity principle has received much less attention than the risk and need principles, and relatively little is known about what can be considered responsivity factors for either men and women; when they are identified, there is often a lack of knowledge on how they influence treatment effectiveness and how to reduce this influence. It is worth noting that most gender-neutral assessments designed to measure justice-involved clients' risk of recidivism and identify correctional treatment needs (e.g., risk/need assessments such as the Level of Service Inventory-Revised [LSI-R]; Andrews & Bonta, 1995) have been developed on samples of men under the assumption that the same risk factors apply to the same degree, and in the same way, for women. Most risk/needs assessments available today are considered gender-neutral. Note that in the current literature review, risk factors and risk assessments that have either a) been developed and used exclusively with samples of men or b) have been developed exclusively on samples of men and then have subsequently adequately validated with samples of women, will be referred to as gender-neutral.

### **Mental Health as a Specific Responsivity Factor**

Despite the relative lack of attention to mental health problems as a risk factor for criminal behaviour in the traditional gender-neutral literature (McCormick et al., 2015), mental disorders have nevertheless been considered a) a specific responsivity issue that has an impact on the effectiveness of correctional treatments, or b) through a broader view of responsivity to impact correctional outcomes indirectly through its impact on other criminogenic needs. This section will discuss mental health as a specific responsivity issue in the context of treatment, and the following section will briefly present evidence pertaining to indirect relationships between mental health and recidivism (i.e., broad responsivity).

### **Empirical Findings**

A diagnosed mental disorder or non-diagnostic indicators of a mental health problem may constitute obstacles to clients' ability to engage in and benefit optimally from correctional programming aimed at reducing recidivism (Bonta et al., 2014). Thus, identifying, attending to, and providing treatment for mental disorders may be important in terms of facilitating clients' successful reintegration into the community following release. The proposition that mental health challenges could hinder treatment progress is an intuitive one. Individuals suffering with major depression may lack the motivation and stamina to complete the program. Others suffering anxiety disorders such as PTSD may be overwhelmed by physical and mental sensations of fear and panic, along with flashbacks to severely traumatic past events, may be incapable of identifying even the most basic program elements. In addition, one would imagine that living in a reality consisting of constant paranoid delusions, intrusive internal "voices" (auditory hallucinations), and surreal physical sensations (i.e., living with a psychotic disorder) would not only make it hard to attend to and understand the program material, but also nearly impossible to understand its' relevance. ADHD is another condition that would be highly disruptive to learning, as it is characterized with a severely reduced ability to maintain attention to an idea, understanding the relationship between distinct ideas and concentrating on educational content, as well as short- and long-term memory deficits.

Despite this intuitively appealing link between mental disorders and program failure, very little empirical research has been conducted to confirm it (McCormick et al., 2015). Nevertheless, some descriptive studies have been conducted to investigate mental health as a specific responsiveness issue. McCormick and colleagues (2017) investigated the relationship between mental health needs, criminogenic needs treatment, and subsequent recidivism in a sample of 232 adjudicated youth (81% boys) on probation in Canada. Mental health needs were

evaluated both in terms of diagnoses as well as elevated but sub-clinical scores (i.e., high scores that fell short of meeting strict diagnostic criteria) using well-validated assessment tools. Youth were identified as having a mental health problem if they had been judged by a clinician to have a mental health disorder or problem and/or to be in need of an mental health related intervention. In this study consisting primarily of boys, mental health needs were not found to be related to recidivism outcomes. However, receiving treatment for mental health needs was associated with also receiving traditional correctional treatment for a larger proportion of criminogenic needs (i.e., youth who received help with their mental health challenges were more likely to receive help with the majority of their criminogenic needs as well (Cohen's  $d = 1.31$ , 95% CI [1.27, 1.34])). This finding may indicate that mental health treatment enabled youth to engage more fully with criminogenic needs treatment. This would support the contention that mental health is a specific responsivity issue. However, this interpretation must be considered cautiously, because it is also possible that the mentally disordered youths were flagged during the assessment and case management planning phases as needing multiple, multi-faceted interventions, explaining why they ended up receiving more comprehensive correctional treatment than youths who were not flagged.

### **Broad Responsivity Factors: Potential Indirect Effects**

Responsivity factors can also be more broadly defined to include those factors that impact criminogenic needs rather than treatment success, and thus indirectly affect recidivism. When responsivity is conceptualized this way, some additional research is relevant. Some studies have found support for this more indirect relationship between mental disorders and recidivism. For example, Visher and Bakken (2014) found that women offenders suffering from a mental disorder were more likely to struggle to obtain stable housing and employment upon release than

women without a diagnosis of a mental disorder. Unstable housing and lack of employment are factors associated with increased recidivism risk. In this way, mental disorder may play a role in increasing these women's failure rates (Van Voorhis et al., 2010; Visher & Bakken, 2014).

In conclusion, within the gender-neutral perspective, mental health problems may be considered relevant to correctional outcomes in one of two ways: a) they may impact the probability of successful correctional treatment, or b) they may reduce women's ability to successfully minimize other risk factors- such as poor accommodations and lack of employment- following release into the community.

### **The Gender-Responsive Perspective**

In sharp contrast to the gender-neutral perspective, the gender-responsive approach emphasizes potential gender differences in the etiology of criminal behaviour, specific risk factors, and treatment needs. It is rooted in early feminist correctional theories (e.g., Chesney-Lind, 1989), which considered gender-neutral theories of crime as inadequate in accounting for women's criminal behaviour. Contemporary gender-responsive paradigms rely on trauma theory, holistic addictions theory, relational theory, and social capital theory (Brown et al., 2020) to explain criminal behaviour among women. The focus in the current context will be on the first three – as they relate directly to mental disorder. These theories posit that trauma, particularly childhood maltreatment, relational dysfunction, and mental health problems can lead to maladaptive behaviours such as substance abuse and in turn dysfunctional survival strategies such as engagement in crime. Both trauma and holistic addictions theory emphasize that substance abuse is a coping mechanism to deal with distress caused by past and current trauma and the disruptive, painful symptoms of mental disorders.

Relational theory emphasizes the importance of relationships-particularly early relationships with caregivers, which become models for adult relationships, as well as romantic relationships, often with antisocial, criminal partners. The dysfunctional relationships are oftentimes characterized by abuse, neglect, unstable and unsafe homes that disturb or interrupt normal developmental processes, including the development of healthy self-concepts, stable identities independent of others, and disturbed notions of mutuality, healthy boundaries, and intimacy. These deficits are later reflected in poor self-esteem and low self-efficacy. Adult relationships may follow dysfunctional childhood models, and can involve co-dependent romantic relations with substance abusing and criminally involved partners; womens' subsequent participation in their partners' activities are frequently thought to be motivated by a desire for connection and closeness. Mutual substance dependence, domestic violence, and other forms of abuse may occur.

In summary, contemporary gender-responsive theories of crime generally emphasize that criminal behaviour, and other life outcomes such as mental disorder and related externalizing behaviours such as self-harm, are attributable to traumatic life experiences. The focus on interpersonal trauma in gender-responsive models is highly consistent with recent research on the effects of adverse childhood experiences, both in the general population and in the population of justice-involved women. It also falls in line with recent developments in psychiatric research and clinical practice regarding the condition known as complex trauma. The complex trauma model posits that the experience of persistent, varied, and severe trauma starting in childhood (e.g., sexual abuse, family violence) and continuing into adulthood (substance abuse, interpersonal violence, domestic assault) have a compounding effect that leads to severe dysfunctions in multiple areas of life. Complex trauma, consequently, is characterized by a plethora of mental

health problems such as internalizing mental disorders (e.g., depression and anxiety), PTSD, ADHD, personality disorders such as BPD, and externalizing problems (e.g., self-harm, suicidal tendencies and interpersonal exploitation and aggression; van der Kolk, 2005). Further, complex trauma is associated with poor self-concept (i.e., unstable sense of self), relational dysfunction (e.g., unhealthy romantic relationships), poor impulse control, reduced emotional regulation that often leads to reliance on substance use to manage emotional pain), and certainly involvement in antisocial activities including substance use, and criminal behaviour such as drug dealing, property crimes, and prostitution).

Gender-responsive scholars vigorously criticize gender-neutral models of criminal behaviour for their lack of attention to women offenders (e.g., inadequate validation of existing gender-neutral risk/need assessments with women offenders). These scholars argue that a) there are specific risk/need factors that are relevant specifically to women (Van Voorhis et al., 2010), b) new assessments for women must include gender-specific factors (Salisbury et al., 2009; Van Voorhis et al., 2010), and c) specialized correctional treatment modalities are needed to adequately address women's needs and effectively reduce their risk of recidivism (Hannah-Moffat, 2009; Van Voorhis et al., 2010; Wright et al., 2012)

### **Gender-responsive risk factors**

There is a limited number of gender-responsive risk/need factors that have been vigorously validated and included in gender responsive risk/need assessments. Among these are adverse (i.e., traumatic) childhood experiences (e.g., physical and emotional abuse and neglect), exposure to trauma in adulthood (e.g., domestic violence, sexual assault), dysfunctional romantic relationships, parenting problems, low self-esteem, low self-efficacy, and notably, mental health problems (Ennis et al., 2013; Van Voorhis et al., 2010). For example, researchers have noted that

justice-involved women have elevated rates of depression and other disorders, that several disorders seem to characterize the profiles of many women, and that a non-negligible portion of empirical studies conducted to date support the contention that some of these disorders are associated with increased risk of recidivism (e.g., Benda, 2005; Van Voorhis et al., 2010).

It is important to note that despite evidence from qualitative pathways studies, mental health, is that despite early qualitative work on offender pathways, and many other approaches to research linking specific mental disorders to the initiation and maintenance of crime remain a-theoretical. Chapter 2 below will provide an overview of theoretical accounts and empirical evidence that provide some support for the potential relationship between mental health problems and recidivism among adult justice-involved women.

### **Study Objectives**

The main objective of the current program of study is to evaluate whether the totality of the resulting empirical evidence supports the view of the gender-responsive perspective that mental disorder and mental health problems may constitute an important risk/need factor among justice-involved adult women. This question will be addressed in two distinct, complementary studies.

The first study (Study 1) is a quantitative meta-analytic summary of the existing empirical evidence linking mental disorder and recidivism among justice-involved adult women. It's main aims were to help determine, based on existing data, what mental disorders and indicators, if any, are risk-relevant and to what degree.

Study 2 was a latent class analysis (LCA) based on a sample of 920 incarcerated women. The major aim of this study was to determine the association between women's latent mental health profiles and subsequent recidivism. This aim encapsulated two sub-goals: 1) the

identification of distinct classes of women who differed in terms of mental health status, and 2) to determine whether those class differences were associated with the hypothesized corresponding variability in recidivism rates.

Chapter 2 starts by providing a comprehensive review of the empirical literature on the connections between the initiation and maintenance of criminal behaviour and mental health disorders and indicators germane to the current meta-analysis (Study 1). A detailed account of the methods employed to prepare for and conduct the analyses follows. Next, the results of the analyses proper are systematically presented. Lastly, a brief discussion of the interpretations and potential implications of the findings and the limitations associated with the study are provided.

Chapter 3 first provides an introduction to the general principles and functions of an LCA, followed by a description of the methods selected for the current LCA. Second, the planned preliminary and main analytic approaches are explained, before the findings of preliminary, main, and post-hoc analyses are sequentially detailed. Lastly, a brief discussion of the substantive interpretation of the findings, their potential implications, and the study limitations are provided.

The fourth and final portion of the document (Chapter 4) is dedicated to bridging the complementary studies' results with the aim of illuminating the overarching conclusions that can be drawn from them. The findings are evaluated in terms of where to best situate them within the gender-neutral vs. gender-responsive debate and how they may help direct future research. Finally, recommendations for risk/need assessments for justice-involved adult women and correctional practice with this population in general are cautiously suggested.

## **Chapter 2**

### **Mental Health and Adult Women's Recidivism: A Meta-Analysis**

#### **A Brief Overview of the Literature**

Although the potential direct associations between various mental disorders and recidivism have been investigated in a few empirical studies, contemporary, comprehensive theoretical accounts are almost non-existent. Nevertheless, preliminary theoretical models have been proposed for a limited number of symptom clusters and sparse, but valuable empirical studies have investigated some of these perspectives. In addition, a minority of symptoms can be said to be intuitively risk-relevant. Below follows a brief review of a selection of symptom clusters theoretically and empirically linked with criminal behaviour. Disorders most commonly known to involve these clusters are highlighted and briefly explored. Following this exploratory section, high-quality empirical studies that further illuminate the initial associations between symptoms, disorders, and criminal behaviour are presented. Note that the current review will not address APD or substance abuse as these disorders have undergone thorough empirical scrutiny and are already included in almost all contemporary risk/need assessments (e.g., LSI-R; Andrews & Bonta, 1995).

#### **Symptom Clusters, Mental Disorders, and Theories of Risk Relevance**

##### **Disinhibition**

Disinhibition in the form of impulsivity, low self-control, short-term non-consequential thinking, and fearlessness may be related to criminal behaviour. In ADHD, impulsivity is a key diagnostic criterion; theoretical accounts have relied on this symptom to explain the well-established empirical link between ADHD and antisocial behaviour (theoretical models: Moffitt, 1993; Sampson & Laub, 1993; empirical studies: Pratt et al., 2002; Rabiner et al., 2005).

Impulsivity is included either directly or indirectly (e.g., via the antisociality construct) in current risk/need assessments. BPD and bipolar mania also share impulsivity as a common trait, and individuals suffering either condition evidence disinhibited expressions of emotions and reduced inhibition in terms of violating social expectations. Unpredictable, potentially dangerous behaviours with severe negative long-term consequences are not unusual.

### **Hyperarousal**

Chronic hyperarousal, hypervigilance, increased attention to threat cues, and severely dysregulated emotions are symptoms closely associated with stress and trauma-related conditions, particularly PTSD. Theorists propose that constant perceptions of potentially lethal threats in the environment and a chronically over-aroused central nervous system can lead to severe anger resulting in violent crime (Marshall, 2002). The proposed link between PTSD and violent crime is supported by empirical evidence (Kulka et al., 1990). Beyond a simple increased violence potential, one may speculate that a sense of foreshortened future could make long-term consequences of, for example, crime, seem irrelevant, and reduced fear of consequences so as to allow individuals to engage in many forms of criminal behaviour.

### **Negative Emotionality and Aggressiveness**

Under-controlled anger, hostility, chronic embitterment and grudges, and irritability occur in many disorders including major depressive disorder (MDD) and PTSD. These symptoms may also occur in sufferers of BPD and NPD. In these disorders however, elevated trait aggressiveness (BPD; NPD), intense self-absorption (BPD), a grandiose sense of self-importance (NPD), and an unusually strong sense of entitlement (NPD) further contribute to the already risky dysregulated negative emotionality. Theorists have proposed that the empirical association between BPD and violent crimes (Moore et al., 2017), including domestic violence

(Jackson et al., 2015) result from both dysregulated anger (Jackson et al., 2015), and elevated trait aggressiveness (Lowenstein et al., 2016; Moore et al., 2017). Acts of violence among individuals diagnosed with MDD and PTSD are thought likely to be due to under-controlled anger and, as mentioned above, fear.

### **Delusions and Hallucinations**

Persecutory delusions, paranoid ideation, chronic suspiciousness and distrust-sometimes coupled with delusions of grandeur- can invoke both fear and anger, potentially resulting in violent self-protection efforts. These symptoms are most often seen in psychotic disorders like Schizophrenia, but can also occur in other mental conditions such as delirium, bipolar mania, and PTSD. Hallucinations are among diagnostic criteria for Schizophrenia and other forms of psychoses, and can further add to fear, rage, and hopelessness. Further, hallucinations can involve false auditory perceptions of authority figures compelling the individual to take certain actions – in some cases including serious acts of unprovoked violence. Empirical research supports delusions and command hallucinations as risk factors for violence (Lamsma & Harte, 2015). Of note, part of the typical progression of many of these disorders is a striking detachment from social norms, which can result in category B sex offenses such as public urination or nudity and other public disturbances constituting crimes. It is crucial to note however, that risk of criminal behaviour rarely persist in psychotic patients following psychosocial and/or medical intervention; the symptoms are only present and relevant during episodes of active psychosis.

### **Risk Seeking**

Common manifestations of thrill-seeking include gambling, illicit sex and prostitution, attending of illegal establishments, driving under the influence, reckless driving, instigation of public fighting, and drug-related offenses. Deliberate and serious risk taking and antisocial

behaviour in general is part of the diagnosis of BPD. However, the sensation-seeking aspect of Bipolar mania also fits this symptom grouping, especially as it co-occurs with the over-confidence and perceptions of overly positive future prospects associated with bipolar manic episodes.

### **Responsivity: Revisiting the Risk-relevance of Mental Disorder in Treatment Contexts**

The majority of studies examining the potential relationship between mental health problems and recidivism have been conducted with samples of men. As discussed in Chapter 1 above, the gender-neutral perspective has long dominated the risk/need correctional literature. Recall that in the gender-neutral perspective, factors that are risk-relevant and targeted in correctional interventions for men are assumed to be relevant in the same way, to the same degree, and for the same reasons for women. The current program of study is meant to shed light on the validity of this assumption and thus a brief review of findings with justice-involved men is worth undertaking to facilitate a comparison with findings obtained with samples of women. Thus, this section will provide a brief example of a comprehensive study of the relationship between mental health and recidivism among men, followed by a review of findings obtained with women,

A number of individual studies of men are available. However, given that the purpose of this brief summary is to provide one point of comparison with findings from studies of women, it is limited to the most comprehensive recent study involving men- a meta-analysis by Bonta et al. (2014). Bonta and colleagues (2014) examined the predictive validity of eight clinical psychiatric indicators (psychosis, mood disorders, personality disorders (PD; Any), APD/psychopathy, intellectual impairment, prior psychiatric admissions, length of inpatient stay in psychiatric hospital, and psychiatric treatment history) in a sample of mentally disordered offenders

(MDOs). MDOs are those offenders who are declared to be mentally disordered during the court proceedings for their index offense, and subsequently housed in forensic mental health hospitals (FMHHs). General and violent re-arrests, re-convictions, or returns to closed psychiatric facility for a new offense were assessed. The meta-analysis included 126 studies ( $N = 23,900$ ). The sample was predominantly male, but also included some women.

Findings of the random effects model<sup>2</sup> indicated that only two clinical variables predicted general and violent recidivism: Any (unspecified) PD and APD/Psychopathy were found to be moderate predictors of general recidivism ( $d = 0.44$  and  $d = 0.54$ , respectively) and violent recidivism ( $d = 0.41$  to  $d = 0.66$ , respectively). A comparison of MDOs, non-mentally disordered offenders (NMDOs) and individuals found not criminally responsible on account of mental disorders (NGRIs) revealed no group differences. Moderator analysis revealed that psychosis was associated with a very modest increase in general and violent recidivism ( $d = 0.21$  to  $d = 0.18$ , respectively) among MDOs and NGRIs.

The meta-analysis had limitations. First, when MDOs with psychosis were compared to MDOs without psychosis, there were no efforts made to ensure group equivalency on other mental disorders. Thus, either group could have had elevated rates of one of the other disorders examined, making it impossible to decipher what disorder actually most influenced the resulting effect size and its significance.

Second, the “healthy” comparison group (NMDO) was simply a sample of justice-involved men (and some women) who had not initially been labelled MDO by the courts. Within

---

<sup>2</sup> The random effects model assumes that there is true variability between studies over and above measurement error, which is preferable to the fixed model given that studies in this literature vary greatly in terms of sample, methodology, measurement, predictor, and outcome definitions, and many other factors that can affect the resulting effect size and confidence intervals.

the NMDO population, a majority of individuals nevertheless have high rates of diagnosed mental disorders (refer back to Table 1 in Chapter 1). Thus, a large portion of the NMDOs also consisted of mentally disordered individuals. To be clear, the authors' comparison involved a sample consisting entirely of all MDOs and a sample wherein a large proportion also suffered mental disorder. This certainly obscures the relationship between mental disorder and recidivism in this analysis.

Third, the authors did not disaggregate findings by gender or indeed utilize gender as a moderator in any analyses. Thus, it cannot be determined whether or how these results were affected by the inclusion of women, or whether they are replicable in all-women samples.

The remaining limitations of the study are provided here, but will not be discussed at length: 1) no differentiation between lifetime and current diagnoses was made. It seems plausible that only currently symptomatic patients would have elevated recidivism rates, 2) treatment and medication adherence were ignored. It seems plausible that if the MDOs received any kind of medical and/or psychosocial treatment, which appears likely in a forensic hospital setting, initially risk-relevant disorders might appear entirely unrelated or very weakly related to recidivism, and 3) the mood disorders category included both major depressive disorder (MDD) and bipolar disorder (BD). Examinations of their respective symptom clusters (see Chapter 1) and existing empirical findings, suggest depressive disorders may be associated with very modest, often non-significant increases in recidivism. Conversely, bipolar disorder has consistently been found to significantly predict of recidivism. Future studies should take care in separating disorders which symptom clusters intuitively suggest differential relationships with antisocial and criminal outcomes. Lastly, the authors did not address the fact that personality disorders of the antisocial cluster: BPD, NPD, APD, and Psychopathy are partially diagnosed

using previous antisocial behaviour. For example, the Psychopathy Checklist Revised (PCL-R; Hare, 1998) – a scale among the most widely used psychopathy measures in North America – includes one subscale – Antisocial – that primarily consists of items reflecting previous antisocial behaviour such as juvenile delinquency, revocation of conditional release, and criminal versatility. Moreover, in a typical assessment, all items on this scale are informed partially or entirely by analyses of the nature of the crimes resulting in convictions. For example, a man convicted of fraud will generally receive a high score on items such as pathological lying and conning/manipulative). Two important points can be taken away from this. First, psychopathy and to some extent APD diagnoses lack sound etiological bases and involve circular rationalizations as part of assessment: Client A commits violent crimes because he has a psychopathic personality disorder. Client A has a psychopathic personality disorder because he commits violent crimes. Second, an effect size representing the relationship between a PD along the antisocial continuum and recidivism partially represents the simple relationship between criminal history and recidivism: past behaviour predicts future behaviour. Findings of studies examining whether the affective, cognitive, and interpersonal characteristics of these disorders contribute meaningfully to risk prediction above and beyond antisocial behavior, and if so to what extent, are inconsistent, with a non-negligible number reporting null findings (e.g., Olver & Wong, 2015) found no significant relationship between scores on affective and interpersonal items (i.e., Factor 1) on the PCL-R and for any recidivism outcome). In conclusion, it is not yet clear whether the so-called internalizing symptoms of antisocial personality styles, actually have a clinically meaningful causal effect on antisocial behaviour among those diagnosed.

In conclusion, the most comprehensive, male-centered meta-analysis to date suggests that only PDs are related to recidivism. However, given the study's limitations, further research is

needed. A meta-analysis focused exclusively on justice-involved women should also be prioritized, as none have been conducted thus far. The next section provides some insight into results obtained in individual studies with all female samples.

### **Women's Mental health and Recidivism**

There is a dearth of research examining the potential relationship between mental health problems and recidivism among adult women offenders (King et al., 2018). No meta-analysis of the relationship between mental disorders and recidivism has yet been conducted for justice-involved adult women. The available individual empirical studies have yielded inconsistent results, with some finding that mental disorders are associated with increased recidivism (e.g., Benda, 2005; King et al., 2018), whereas others have found little or no support for such a link (e.g., Blanchette, 1996; Blanchette & Motiuk, 1996; Du et al., 2013). Rather than providing a summary of the available literature in its entirety, this section will present a few recent empirical investigations that involve a range of sample types, settings, mental health and recidivism variables, follow-up periods, methodologies, and analytic strategies, and that aid in highlighting typical inconsistencies in findings both within and across studies.

Between the years 2008 and 2010, across four distinct studies, Van Voorhis and colleagues' research group have arguably undertaken the most extensive and systematic investigation of potential gender-responsive risk factors for recidivism, including mental health variables, to date. This section will provide a summary of overall findings from these four studies. The findings reported here pertain to those derived from univariate analyses involving dichotomous recidivism outcomes. A large number of samples were utilized. One sample of probationers in Maui ( $N = 158$ ) was examined. Six (five independent) Missouri samples: probationers (two partially overlapping ( $N = 272-298$ ) and one independent sample ( $N = 304$ )),

two pre-release ( $N = 150$ ;  $N = 272$ ), and one prison ( $N = 244$ ) sample. Recidivism was measured as breaches, any returns to custody, and rearrests. Six, 12, and 24 months follow-up periods were examined. The mental health variables examined were: 1) Any lifetime diagnosis of mental disorder (MD), 2) current depression/anxiety, 3) current psychosis/suicidality, and 4) current psychosis. All but Any MD were continuous total scale scores or symptoms. Any lifetime diagnosis of a mental disorder was explored only in the Maui probation sample ( $N = 158$ ). Having a lifetime diagnosis of any mental disorder did not significantly predict recidivism<sup>3</sup>. Depression/anxiety were predictive in some samples but not others. For all probation samples (four), elevated anxiety/depression significantly but modestly predicted breaches and returns across all follow-up times ( $r = .12$  to  $r = .20$ ; See Table 2 below)<sup>4</sup>. Conversely, in the prison ( $N = 244$ ) and pre-release samples ( $N = 150$ ;  $N = 272$ ), no relationship between depression/anxiety and recidivism was found.

The psychosis/suicidality predictor was only assessed in the probation sample of 298 women and only for breaches at six and 12-months follow-up: Current psychosis/suicidality predicted recidivism modestly but significantly ( $r = .11$  to  $r = .12$ ,  $p < .05$ ; see Table 2 below for a summary of findings). Psychosis also yielded inconsistent findings. In the two pre-release samples, higher psychosis scores were modestly but significantly associated with increased rates of returns at 12 to 24 months follow-up (see Table 2 below for a summary of results).

---

<sup>3</sup> There was a single exception for 24 months returns to custody, for which a very small but positive and significant effect size resulted ( $r = .13$ ,  $p < .01$ ).

<sup>4</sup> Only one of the total of 10 effect sizes reached was non-significant, reaching only marginal significance ( $r = .09$ ,  $p < .10$ ). This occurred in the probation sample consisting of 304 women for technical violations at six-months follow-up.

**Table 2**

*Mental Health and Recidivism Among Adult Justice-Involved Women: Overview of Selected Studies and Findings*

Study	N	Mental health indicators	Follow-up	Recidivism Predictor
Becker et al. (2011)	1, 542	- 3+ inpatient/emergency care admissions - 1+ involuntary psychiatric examinations - 12+ outpatient mental health contacts	48 months	<b>Mixed</b> Yes No No
Benda (2005)	300	- change in stress - change in depression - change in fearfulness - change in suicidal thoughts	60 months	<b>Mixed</b> Yes Yes Yes No
Blanchette (1996) <sup>a</sup>	66	- MMD - GAD - MDD - Psychosexual dysfunction - APD - Substance/alcohol dependence	24 months	<b>No (mixed)</b> No No; inversely related <sup>c</sup> No No Yes Yes
King et al. (2018)	2, 311	- SMD <sup>b</sup>	96 months	<b>Yes</b>
Van Voorhis et al. (2008/10) <sup>b</sup>	1, 426	- Any lifetime MD - Current anxiety/depression - Current psychosis/suicidality - Current psychosis	Six months 12 months 24 months	<b>Mixed</b>  <b>No</b> <b>Mixed</b> <b>Yes</b> <b>Mixed</b>

*Note.* MMD = Major Mental Disorder (schizophrenia, schizophreniform disorder, mania); GAD = Generalized anxiety disorder; MDD = Major depressive disorder; APD = Antisocial personality disorder; SMD = Depressive disorders, Schizoaffective disorder, Bipolar disorder, Dissociative Disorders; MD = Mental Disorder. <sup>a</sup> Overlaps with Blanchette & Motiuk (1996), which was also used.

<sup>b</sup> Results were sample-dependent. <sup>c</sup> Women with GAD had non-significantly lower recidivism rates

In summary, 90% of correlations for depression/anxiety, 67% of correlations for psychosis/suicidality, and only 36% of correlations for psychosis supported these mental health variables as valid, albeit modest, predictors of recidivism. Inconsistencies may have several explanations. First, the averages, range, and distributions of scores may have differed across sites (e.g., depression/anxiety scores may have been more uniformly high in prison relative to in probation samples). Second, there was no attempt to match samples on other important variables a-priori. Thus, samples could have differed systematically in terms of mental health profiles, correctional profiles, estimated risk, and criminogenic needs, as well as other relevant factors. Finally, the authors did not mention whether the scales and scoring procedures remained static across sites; the possibility remains that the measures were not directly comparable across samples.

In a large study from North Carolina, King and colleagues (2018) investigated the relationship between severe mental disorders (SMD) and recidivism in a sample of 2,311 incarcerated women. SMD included depressive disorders, bipolar disorder, schizoaffective disorder, and dissociative disorder. Twenty percent of the sample ( $n = 453$ ) were diagnosed with an SMD. Recidivism was defined as a return to custody in a state prison in North Carolina and the length of follow-up was eight years. The authors did not report whether returns were restricted to new offenses or included technical violations (i.e., breaches). In logistic regression analyses using a binary recidivism outcome (*yes vs. no*), risk relevant covariates were controlled for; age, race, marital status, number of children and prior incarcerations, and substance abuse. SMD did not significantly predict recidivism ( $OR = 1.2$ , 95% CI = [0.96, 1.51]  $p = .11$ ). In addition, a time-to-recidivism cox regression analysis was conducted. The hazard ratio indicated that women with SMD were at 16% greater hazard to recidivate than women without SMD ( $p <$

.05). There were two major limitations to the study. First, SMD did not include other conditions that can be very severe, such as BPD, PTSD, or other psychotic disorders such as Schizophrenia. If a meaningful proportion of women in the comparison group suffered any of these disorders<sup>5</sup>, the relationship between SMD and recidivism would likely be attenuated. Further, the researchers did not assess symptom severity, past or current treatment, or medication adherence among SMD women. The possibility thus remains, that some portion of these women were not experiencing severe symptoms during the follow-up period, again potentially resulting in small and insignificant effect sizes (see Table 2 below for a summary of results).

Benda (2005) examined the relationship between *change* in stress, depression, fearfulness, and suicidal thoughts and recidivism-re-arrest for a new offense or breach of parole among 300 women and 288 men sentenced to boot camp in the United States. The length of follow-up was five years. In cox regression analyses of relationships between *change* in continuously measured stress, depression, and fearfulness, but not suicidal thoughts were robust predictors of time-to-recidivism among women. Stress and suicidal thoughts predicted time-to-recidivism for men (see Table 2 below for a summary of results).

The inconsistent findings typical of the literature on mental health and recidivism is evident in Becker and colleagues' (2011) study of 3, 769 individuals who spent one or more days in a county jail in the United States (59% men; 41% women). The only three binary mental health variables included 1) 3+ emergency room/psychiatric hospitalizations, 2) 1+ involuntary psychiatric exams, and 3) 12+ outpatient mental health contacts. Recidivism was defined in

---

<sup>5</sup> Refer to Table 1 in Chapter 1 above for reported rates of other severe mental disorders such as BPD, PTSD, and psychotic disorders. About 30% of incarcerated women suffer BPD. The proportion is comparable for PTSD. Percentages for psychotic disorders may range from about 5-8%.

terms of re-arrests. Individuals with 3+ emergency room visits or hospitalizations were significantly *more likely* to be re-arrested. However, those who had experienced one or more involuntary exams or had 12 or more outpatient mental health contacts, were *less likely* to be re-arrested. A few potential explanatory factors may be relevant in interpreting these results. First, legally enforced psychiatric examinations are highly likely to lead to prescription(s) for medications and can result in in- or –outpatient treatment, reducing or nullifying any initial risk. Second, having multiple outpatient contacts may not adequately measure risk-relevant mental health conditions. This variable may also capture a number of mental health challenges not associated with criminal behaviour. Further, access to multiple formal supports and potentially treatment would be expected to reduce risk. In addition, multiple outpatient contacts may also partially reflect the patient’s willingness to take responsibility for their own mental health, which would be expected to be associated with lower risk. Finally, and most importantly, as mentioned in Chapter 1 above (see Table 1), the rates of mental disorders among incarcerated individuals are extremely high: This means that a very large proportion of the sample may have suffered mental health conditions, but without receiving exams, treatment, medications, or outpatient support. If this were the case, the findings of an inverse relationship between contact with mental health professionals and recidivism would not be surprising (see Table 2 below).

In a study of 66 federally incarcerated women in Canada, Blanchette (1996) and Blanchette and Motiuk (1996)<sup>6</sup> examined the potential association between lifetime diagnoses of major mental disorder (MMD), which included schizophrenia, schizophreniform disorder, and mania (Bipolar disorder), Generalized anxiety disorder (GAD), MDD, and psychosexual

---

<sup>6</sup> Blanchette & Motiuk (1996) is the peer-reviewed publication of Blanchette’s (1996) original unpublished thesis.

dysfunction, as well as APD and substance/alcohol dependence and recidivism. Recidivism events included any return to custody, breaches, and re-convictions. Only diagnoses of APD and substance/alcohol dependence were associated with significantly higher recidivism rates; this finding is not surprising as these disorders are already accepted as major gender-neutral risk factors and are represented in all contemporary risk/need assessments (see Chapter 1 above). A point of interest was the potentially inverse relationship between GAD and recidivism, despite its failure to reach statistical significance ( $r = -.18, p > .05, ns$ ). The negative relationship between GAD and criminal behaviour seems intuitive: chronic anxiety likely constitutes a hindrance to performing risky, adrenaline-filled, sometimes dangerous acts that could easily result in social condemnation, criminal proceedings, and possible incarceration. This possibility of null and negative relationships between some disorders and recidivism is important to highlight; one perpetual methodological flaw of existing studies is the assumption of equivalency between disorders and the resulting tendency to combine potentially incompatible conditions into a single predictor variable in analyses (see Table 2 below).

### **Summary**

The literature speaking to the risk-relevance of mental health for women should be interpreted with caution. The degree of variability in findings both within and between studies points to the need for additional research wherein attempts are made to address the limitations of existing studies. Meeting contemporary methodological standards of observational studies in the fields of psychological and correctional research should be the aim of any investigation. Sound theoretical models accounting for the relationships we hypothesize to find are sorely needed. Further, when group comparisons are made it is necessary to ensure that the “healthy”

comparison group does not consist of women who may also suffer major mental health challenges: a-priori group equivalency on other risk-relevant variables should also be prioritized.

Distinct disorders and indicators should be examined separately, such that the effect of one does not obscure the effect of another. Contextual factors such as treatment status, medication adherence, the current status of the disorder (e.g., active symptoms vs. in remission), and current symptom severity, should be considered whenever possible. Lastly, a largely unexplored possibility is that the relationship between mental health indicators and recidivism depends on the individual woman: there is more than one path to criminal offending (Chesney-Lind & Pasko, 2013). Thus, the establishment of reliable mental health and correctional profiles might also be desirable.

A quantitative meta-analytic review is a necessary first step in the process of investigating and prioritizing justice-involved women. Establishing and empirically validating mental health and overall correctional profiles for women in terms of differences in correctional outcomes would also constitute a necessary step in the process of determining the utility of gender-responsive risk/need factors in risk assessment, treatment decisions, and case management.

### **The Current Meta-Analysis**

The potential association between mental disorders and other indicators of poor mental health and recidivism requires further research (Becker et al., 2011; Blanchette, 1996; Blanchette & Motiuk, 1996; King et al., 2018; Van Voorhis et al., 2010). This study is a meta-analysis of the association between general indicators of mental health problems (e.g., history of psychiatric hospitalizations) as well as specific mental disorders outlined in the DSM-V (APA, 2013), and recidivism among justice-involved adult women.

Despite its limitations, the current-meta-analysis is an important first step towards the longer-term goal of expanding the knowledge and improve the understanding researchers, clinicians, and other stakeholders in the correctional system have of the challenges faced by justice-involved women.

### **Scope of the Meta-Analysis**

The current meta-analysis differs from previous studies and contributes uniquely to the scientific literature in several ways. First, the current meta-analysis will be the first quantitative synthesis of the relationship between mental health and recidivism limited to justice-involved adult *women* to date. Second, both the “white” literature – published, peer-reviewed studies- and the “grey” literature, such as government reports, unpublished dissertations and thesis, and book chapters were searched and studies were included from both. Third, the current analysis includes both prospective and retrospective longitudinal observational studies. Retrospective studies were included only when the temporal order of the mental health assessment and the subsequent time at risk could be absolutely ensured (i.e., cross-sectional studies were excluded). Fourth, the current study’s scope in terms of time period was the most inclusive possible: studies completed from the earliest possible date in individual search engines to the most frequent possible dates were all considered eligible. Fifth, rather than restricting analyses to women who were incarcerated at the time of assessment or documentation (or lack thereof), this study also included samples of women in the community. Importantly, in the current meta-analysis every effort was made to examine one mental disorder at a time, rather than combining various disorders into clusters such as “mood and anxiety disorders” – safeguarding against the potential that the distinct disorders could be differently (e.g., one positively and another inversely associated) related to recidivism. Further limitations of previous meta-analyses involving youth,

such as being limited to broad “catch-all” or even singular global binary variables (e.g., Scott & Brown, 2018), or indeed to the examination only one disorder (e.g., anxiety; Hubbart & Pratt, 2002), were avoided. Further, previous meta-analyses such as Bonta et al. ’s (2014) pioneering study examining the relationship between mental disorder and recidivism among mixed-gender MDOs and NGRIs have been largely limited to those commonly associated with court-determined MDO status (e.g., psychotic disorders, personality styles known to be at increased risk for recidivism [Psychopathy]). In the current study, disorders of interest were not thusly restricted, but rather examinations of more commonly occurring and theoretically relevant diagnoses such as PTSD were permissible. The current meta-analysis attempted to include non-English studies available in one of the languages the current author is proficient; Norwegian, German, Swedish, or Danish. Finally, despite severely limited and not infrequently flawed data, every attempt was made to examine potential moderators in order to provide context to the main findings and thus a better understanding of the conditions under which the observed effects are most relevant.

### **A–Priori Exclusion Criteria**

Of note, despite its very broad scope, the current meta-analysis excluded the personality style APD as well as Substance Abuse Disorders/dependence from quantitative syntheses as 1) their risk-relevance is entirely intuitive, 2) the disorders have been researched at length across populations, 3) they are already considered part of the central eight gender-neutral risk factors (i.e., criminogenic needs), and 3) they part of all contemporary risk/need assessments. Further, effects sizes from analyses with continuous recidivism outcomes (i.e., time-to-recidivism analyses with resultant hazard ratios [*HRs*]) because *HRs* cannot be aggregated with other effect size measures pertaining to binary outcomes.

## **Research Questions**

The main aim of this study was to examine the utility of general mental health indicators as well as specific mental disorders in predicting recidivism among adult justice-involved women. The primary research questions of interest were as follows:

### **1. Are mental disorders predictive of recidivism among justice-involved adult women?**

A-priori, specific disorders and diagnostic categories of interest included 1) depression, major depressive disorder (MDD), and bipolar disorder, and 2) anxiety disorders and post-traumatic stress disorder (PTSD). Anxiety disorders as a diagnostic category and PTSD were considered separate constructs of interest for the following reasons: 1) etiological differences between typical anxiety disorders and PTSD are known to differ supporting the contention that they may relate differently to recidivism, 2) the symptoms of these disorder types differ in clinically meaningful ways supporting distinct expectations regarding their risk-relevance among adult women, and 3) the disorders are assessed differently in both non-forensic psychiatric and correctional contexts (e.g., risk/need assessments; the Service Planning Instrument for Women [SPIn-W; Orbis Partners, 2003]). Further, borderline personality disorder, attention-deficit/hyperactivity disorder, and psychotic disorders (e.g., schizophrenia) were also of particular interest.

### **2. Does the presence of any mental disorder predict general recidivism?**

Any mental disorder as a unitary construct reflecting all relevant mental disorders was to be examined in analyses.

### **3. Do externalizing behaviours indicative of mental health problems such as self-harm and suicide attempts predict general or violent recidivism?**

Past and current non-lethal self-injurious behaviour, suicidal ideation, and suicide attempts were the most relevant externalizing behaviours in the current meta-analysis.

4. **Do general mental health indicators such as psychiatric history recidivism** (e.g., number of previous admissions to psychiatric hospitalizations, psychopharmaceutical medication use, outpatient mental health contacts).
5. **What factors moderate the relationship between mental health disorder and indicators and recidivism?**

### **Hypotheses**

Based on the empirical research literature on the limited theoretical accounts, empirical findings, and critical assessments of potential behavioural outcomes of diagnostic symptom clusters associated with specific mental disorders, and recidivism among adult justice-involved women reviewed above, I made the following a-priori hypotheses:

1. In terms of the specific mental disorders, I hypothesized that depressive disorders were anticipated to very modestly and positively predict general recidivism. Non-PTSD disorders were expected to emerge as a protective factors, negatively predicting recidivism (Blanchette, 1996; Blanchette & Motiuk, 1996; Hubbart & Pratt, 2002; Wibbelink et al., 2017; but see Van Voorhis et al., 2010). PTSD was anticipated to significantly but modestly predict recidivism, particularly violent offenses. Although a violent action-potential can exist for patients suffering severe psychotic disorders in the absence of treatment, I made no firm hypothesis with regards to this diagnostic category. BPD was expected to be moderately and significantly predictive of both general and violent recidivism. Mixed internalizing and externalizing disorders such as ADHD-particularly the hyperactive subtype- as well as

bipolar disorder(s) would be modestly but significantly predictive of recidivism (Becker et al., 2011; Green, 2006; Van Voorhis et al., 2010).

2. The presence of any mental disorder was not expected to be associated with recidivism.
3. Externalizing behaviours indicative of poor mental health such as self-injurious behaviours and attempted suicide were hypothesized to be associated with moderate increases in recidivism rates, particularly violent offenses.
4. Non-diagnostic indicators or mental health such as a history of psychiatric hospitalizations were not expected to be significantly related to recidivism.
5. In term of potential moderating factors, it was expected that binary, formal diagnoses of mental disorders would be more strongly predictive of recidivism relative to continuous scores on measures of mental health, due to the requirement that diagnoses typically require disruption of normal functioning in one or more important areas of life. Further, it was tentatively hypothesized that current mental health assessments would be more strongly associated with all recidivism outcomes than lifetime diagnoses. No hypotheses were made regarding other potential moderators such as peer-review status, length of follow-up, or year of publication.

### **Method**

The main aim of this meta-analysis was to quantitatively synthesize the existing literature regarding the potential association between mental disorders (e.g., ADHD) and recidivism, as well as other general mental health indicators (e.g., self-harm, psychiatric history, etc) and recidivism among justice-involved adult women.

### **Meta-analysis Protocol Registration**

Contemporary standards for the ethical conduction of systematic reviews and meta-analyses include the registration of a complete review protocol prior to study commencement. In other words, researchers are expected to submit their overall research plans and any supporting documentation such as the coding manual to an official register, wherein the submission is reviewed in full, and upon approval, made available to the public. The main aim of this practice is to increase transparency and accountability particularly in the context of the updating of meta-analytic results. The National Institute of Health Research (NIHR; UK), through its Center for Reviews and Dissemination (CRD) at York University, has developed the largest international online record of meta-analyses and reviews: The Prospective Register of Systematic Reviews (PROSPERO), which provides a platform from which researchers can register their reviews and meta-analyses. The recommendations by one of the leading authorities on reporting standards for meta-analyses - the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) – include registration within PROSPERO in their official guidelines. Thus, the current meta-analysis was submitted to the PROSPERO register and has been published there (Record number: CRD42020157668; the date of initial submission receipt was June 17<sup>th</sup>, 2020).

### **Inclusion Criteria**

To be considered for inclusion the studies had to include (a) a sample of adult justice-involved women, b) a minimum total sample size of 10 women, (c) a measure of a mental health disorder or indicator used to assess mental health status in the relevant sample (see Exclusion criterion (a) below for excluded indicator types), d) a outcome measure of recidivism (see Exclusion criteria (d) below), and (d) an effect size or a sufficient amount of data from which an effect size could be calculated.

**Exclusion Criteria**

Excluded from this review were (a) studies utilizing exclusively inappropriately mixed mental health indicators- meaning combinations of variables that may be anticipated to have opposite relationships with recidivism (e.g., the Emotional/Personal Domain on the LSI-R is considered an indicator of mental health, but consists of one score for self-harm, prior treatment, ADHD, depression, and anxiety; Andrews & Bonta, 1995), (b) studies examining exclusively APD, Psychopathy, or substance abuse/dependence disorders, (c) studies for which univariate effect sizes could not be obtained, (d) studies limited to continuous recidivism outcomes and resulting non-convertible Hazard Ratios (*HRs*).

**Additional Post-hoc Exclusion Criteria**

Post-study collection and data extraction, the inclusion criteria were further refined to specify that recidivism measured had to be binary and that only effect sizes from univariate analyses would be included. Samples of women legally classified as mentally disordered by the courts (MDOs) were initially accepted and coded, but later excluded for several reasons. First, of the four MDO studies, two conducted group comparisons wholly inappropriate for the current study (Davis, 2007; Steels et al., 1998). In these studies, the disordered group consisted of individuals diagnosed with schizophrenia and a non-mentally disordered comparison group consisting exclusively of individuals described as non-mentally (i.e., no Axis I disorders) disordered but psychopathic. Analyses consisted of group comparisons and no effect size was provided for a continuous measure of symptoms of schizophrenia and recidivism alone to the exclusion of the psychopathic comparison group. Evidence supporting any association between

schizophrenia and non-violent general recidivism is virtually absent<sup>7</sup>. Justice-involved men and women with psychopathic personalities have been found to have elevated general and violent recidivism rates. Although research on women with psychopathy is limited, some inferences can be drawn based on studies of their male counterparts. Men with psychopathic personalities have been found to have 20 to 25 times the odd of being incarcerated and recidivism rates four to eight times as high as their otherwise comparable non-psychopathic counterparts (Kiehl & Hoffman, 2011). Thus, including these comparisons in the aggregate effect sizes could obscure and attenuate the relationship between mental disorders and recidivism. In another study, only multivariate effects were reported (Lanes, 2005). In the fourth study, only violent recidivism outcomes were examined. Only two studies in the meta-analyses reported effect sizes for violent recidivism and these pertained to different mental health predictors. In order to compute an aggregate effect size using meta-analysis, three is the absolute minimum number of independent effect sizes required (Helmus & Babchishin, 2013). Thus, there were an insufficient total number of studies reporting effects for violence recidivism for it to be examined in the current meta-analysis (Harris, 2002).

### **Literature Search**

A thorough search of both the published and unpublished literature was conducted. Published journal articles, book chapters, unpublished manuscripts, published and unpublished dissertations and theses, and government and university reports were considered for inclusion.

---

<sup>7</sup> Theoretically and empirically, schizophrenia is linked with violent outcomes but only when delusions and command hallucinations are present and left untreated (Keers et al., 2014; Lamsma & Harte, 2015). For patients suffering schizophrenia, involuntary incarceration in forensic mental health hospitals almost invariably involves treatment with antipsychotic medication as well as psychosocial therapies. When medication compliance is achieved, the violence risk believed latent in psychotic disorders is believed to diminish.

The period covered by the search started at the earliest possible date (i.e., start date was left blank) and ended in March 31st of 2020. The Carleton University Library's search engine allows for the search of a multitude of databases simultaneously and was therefore used to efficiently conduct the main literature search of scholarly databases. The following academic databases were considered the most relevant and were included in all searches: Criminal Justice Abstracts, Dissertations and Theses @ Carleton, ProQuest Dissertations & Theses Global, Foreign Doctoral Dissertations, Medline, National Criminal Justice Reference Service (U.S.), Open Access Theses and Dissertations, ProQuest Databases, PsycArticles, PsycBooks, PsycINFO, PubMed, Sociological Abstracts, Social Sciences Fulltext, Theses Canada, and Web of Science. In addition, multidisciplinary databases such as Academic OneFile, Cambridge Journals Online, Cambridge University eBooks, SAGE Journals Online, ScienceDirect Journals, SpringerLink Journals, Taylor and Francis Journals Online, Wiley Online Journals, and Worldcat were also included. Further, a careful search of one specific journal, the Criminal Behavior and Mental Health journal (CBMH) was conducted separately.

Interestingly, the careful review of several initially eligible full text documents retrieved during the search of academic databases cited additional studies available from university websites that should be screened for eligibility. This discovery led to additional searches of a number of academic websites, including the University of Nevada, Las Vegas (UNLV) and the Center for Crime and Justice Policy. The following non-academic and alternative databases were also searched: Australian Institute of Criminology, the National Institute of Corrections (NIC; U.S.), and the National Institute of Justice (NIJ; U.S.). A number of government websites were also searched: Correctional Services of Canada (CSC; Canada), Department of Justice Northern Ireland (NIR), Federal Bureau of Prisons (FBP; U.S.), Her Majesty's Prison Service (HMPS;

U.K.), Ministry of Justice (U.K.), New South Wales Bureau of Crime Statistics and Research (BOCSAR; Australia [AU]), New Zealand Department of Corrections (New Zealand [NZ]), Public Safety Canada (PS; Canada), Queensland Corrective Services (AU), South Australia Correctional Services (AU), Tasmania Corrective Services (AU), Victoria Correctional Services (AU), and Western Australia Corrective Services (AU). Further, Researchgate - a website containing the profiles of a large number of researchers and their work globally- as well as Google Scholar- were searched to ensure that all relevant studies were identified.

Manual searches of the reference lists of already collected documents and relevant previous meta-analyses and review articles were conducted. Further, Blais et al. (2014) published a meta-analysis examining the associations between specific mental health disorders and general indicators and recidivism primarily among MDOs. The studies cited in this meta-analysis were also reviewed. Notably, a number of the cited studies included some proportion of justice-involved women ( $k = 73$ ). Upon inspection of these studies, it became clear that the majority did not aggregate their findings by gender, a number contained too few women to meet the inclusion criterion of at least 10 women participants, and a smaller proportion included only men. When the study did not aggregate the findings by gender but appeared to contain 10 or more women participants, an attempt to contact the lead or the corresponding, author was made in an effort to obtain unpublished women-specific data (see Appendix G for list of authors and results on inquiries).

Authors were also contacted if it was believed that they were in the possession of unpublished raw data or manuscripts, clarification of reported findings was needed in order to determine whether a collected study met the inclusion criteria or not, additional data was required (e.g., to obtain a recidivism base rate that was not reported in the publication), or the

authors only reported multivariate effects, in which case they were contacted in an effort to obtain unpublished univariate data. The results of the attempts to contact the authors are reported alongside the results of the literature review below (see Appendix G as well).

### Search Terms

The bulk of the literature review centered on sources that require English language search terms. These English language search terms defined the population (*women offend\**, *female offend\**, *female inmate\**, *female prisoner\**), general terms for mental disorders (*mental health\**, *psychiatric\**, *mental disorder\**, *mental illness\**), broad categories of disorders (*axis I disorder\**, *axis II disorder\**, *externalizing disorder\**, *internalizing disorder\**), terms frequently used in a number of publications (*severe mental disorder\**, *major mental disorder\**, *severe mental illness\**, *severe psychiatric illness\**, *severe psychiatric disorder\**), and an extensive range of general and specific terms for mental disorders as well as their synonyms and abbreviations included in the DSM-IV (APA, 2000) or the DSM-V (APA, 2013)<sup>89</sup>. Both copies of the DSM were used because older studies likely relied on the fourth-edition. Note that some disorders were excluded from the search as they did not pertain to the current population or context (e.g., *mixed receptive-expressive language disorder* is only relevant prior to adulthood and could not be measured in current assessments of adult women). Mental disorder terms included: (1) neurodevelopmental disorders/disorders first diagnosed in childhood (*developmental disorder\**, *autistic disorder\**, *autistic spectrum disorder\**, *autism*, *Asperger's\**, *attention-deficit/hyperactivity disorder*, *ADHD*, *ADD*, *Tourette's\**, *tic disorder*), (2) neurocognitive

---

<sup>8</sup> Synonyms to the disorders of the DSM-V (APA, 2013) were included as search terms where appropriate.

<sup>9</sup> Substance-related disorders were excluded from the study and the search terms as these have been examined in depth in previous studies (Andrews et al., 2012; Olver et al., 2014)

disorders (*delirium, dementia, amnesic disorder, cognitive disorder\**), (3) psychotic disorders (*psychos\*, psychotic\*, Schizo\*, delusional disorder, paraphrenia*), (4) mood disorders (*mood disorder\*, depress\*, major depressive disorder, major depression, dysthym\*, persistent depressive disorder*), (5) bipolar and related disorders (*bipolar\*, bipolar I, bipolar II, bipolar not otherwise specified [NOS], bipolar disorder NOS, manic-depression, mania, manic\*, cyclothym\*, major depressive disorder*), (6) anxiety disorders (*anxiety\*, generalized anxiety\*, GAD, panic attack\*, panic disorder, agoraphobia, specific phobia\*, phobia, phobic, social anxiety disorder, separation anxiety disorder*), (7) obsessive compulsive and related disorders (*obsessive-compulsive\*, OCD*) (8) trauma- and stressor related disorders (*post-traumatic stress disorder, PTSD, acute stress disorder, adjustment Disorder*), (9) somatic symptoms or related disorders/somatiform disorders (*somatiform disorder, somatic symptom disorder, somatization, conversion disorder, pain disorder*), (10) *factitious disorder*, (11) dissociative disorders (*dissociat\*, depersonalization\**), (12) sexual and gender-identity disorders (*gender-identity disorder\*, paraphilia, exhibitionism, frotteurism, pedophilia, masochism, sadism, transvestic\*, voyeurism*), (13) eating disorders (*eating disorder\*, anorexia\*, bulimia\*, binge eating\**), (14) sleep disorders (*sleep disorder\*, insomnia, sleep walking disorder*), (15) disruptive, impulse-control, and conduct disorders/ impulse-control disorders not elsewhere specified (*impulsive control \*, intermittent explosive\*, kleptomania, pyromania, pathological gambling, trichotillomania*), and (16) personality disorders<sup>10</sup> (*personality disorder\*, paranoid personality\*, PPD, borderline personality\*, BPD, narcissistic personality\*, NPD, obsessive-*

---

<sup>10</sup> Antisocial personality disorder was excluded from the study and the search terms as a proxy for this disorder has already been examined in depth in previous studies (i.e., studies including the antisocial personality pattern of the LSI-R; Andrews et al., 2012; Olver et al., 2014).

*compulsive personality\**), some of the commonly reported non-diagnostic mental health indicators, including self-injurious behaviour (*self-harm\**, *self-injur\**, *suicide\**, *suicidal ideation\**, *suicidal thought\**, treatment history (*psychiatric hospitalization\**, *psychiatric treatment\**, *in-patient\**, *out-patient\**, *involuntary committal\**, *mental health treatment\**), psychotropic medication use (*psychotropic\**, *psychotropic medication\**, *psychopharmaceutical\**), and finally, the outcome of interest-recidivism (*recidivis\**, *reoffen\**, *new charge\**, *re-offen\**, *re-arrest\**, *rearrest\**, *reconvict\**, *re-conconvict\**, *reincarcerat\**, *revocation\**, *technical violation\**, *breach\**, *returns\**, *return to custody*, *return to prison*).

A number of non-English sources were also searched for eligible studies, using both English and non-English search terms. The sources and terms are available in Appendix A. Notably, no additional studies were identified using these non-English sources or terms. The Literature Review Results section below displays three separate figures (Figures 2, 3, and 4), outlining the search and screening procedure separately for academic databases (Figure 2), alternative non-academic database sources (Figure 3), and a total overview of the studies included in the meta (Figure 4) across sources. The results section also outlines search strategy modifications that had to be made to effectively utilize particular sources, as well as details manual searches and results of attempts to contact study authors to obtain additional data or clarification of findings.

### **Coding**

The coding manual, including all instructions as well as the electronic coding sheets, is provided in Appendix A. A number of strategies were employed in order to identify the variables that should be extracted from each study. Initially, I used five heuristics to determine the variables to be extracted and coded. First, I derived a number of variables that, based on the

published academic literature on mental disorders, justice-involved women, and criminal behaviour generally, might plausibly influence (e.g., moderate) the observed association between mental disorders or indicators and recidivism. For example, both the timing (e.g., *lifetime* vs. *current* diagnoses), and the nature (*dichotomous* vs. *continuous*) of measurement of mental disorder could potentially influence the resulting effect size. Second, I carefully reviewed the variables identified in recent (i.e., 2010 and later), high-quality meta-analyses published in well-established academic journals (e.g., the British Journal of Psychiatry; 2018; Clarivate Analytics, 2019). Among others, I reviewed Bonta et al. (2014), Fazel and Seewald (2012), O’Shea and Dickens (2016), Olver et al. (2014), Scott and Brown (2018), Young et al. (2014), and Wibbelink et al. (2017). Third, following their review of my initial coding manual, advice regarding potential additional variables that should be extracted, was provided to me directly from three experienced researchers and professors at the Department of Psychology at Carleton University, including Dr. Blais, who co-authored the Bonta et al. (2014) meta-analysis, Dr. Serin whose research area includes decision making and correctional programming, and Dr. Brown- an expert in the area of gender and crime. Fourth, at multiple times throughout the process of screening, evaluating, and coding of the eligible studies, promising additional variables were identified and added to the existing coding manual. Previously completed coding forms for each study were then continuously updated with each new addition. Finally, once all included studies had been coded and a dataset had been compiled to accommodate the necessary quantitative synthesis, yet another a small number of additional variables were added. These variables were modified versions of already coded variables (e.g., based on the item *average risk level* an item entitled *majority risk level* and another *percentage belonging to various risk classifications* were created. This was to avoid a situation wherein, after coding of many studies, it was discovered that

average risk level was almost never reported while majority risk level was. By coding both simultaneously, returning to studies already coded to search for additional data would not be necessary. The entire coding manual and associated blank coding sheets are presented in Appendix A.

### **The Complete Coding Protocol: Overall Structure and Main Features**

The final complete coding manual (Appendix A) consists of four main sections. Section one pertains to study characteristics, such as publication status (*published vs. unpublished*), peer review status, (*yes or no*), and study design (*retrospective vs. prospective*). Section two allows for the coding of sample characteristic for the overall samples before and after attrition, as well as any number of sub-samples. Example items include (sub)sample size(s), the predominant mental disorder present in the sample, and mean age.<sup>11</sup> The third section of the coding protocol pertains to the coding of effect size data for group comparisons. The variables included for group comparisons were the following: mental health diagnosis or indicator (i.e., type of mental health problem assessed), measurement of the mental health predictor (i.e., *dichotomous vs. continuous*), *current vs. lifetime* diagnosis, assessment method (e.g., *self-report vs. clinical interview*), assessment protocol (i.e., if *clinical interview* was used), *name of assessment measure*, recidivism information (i.e., *type of recidivism*), definition of recidivism (e.g., *re-*

---

<sup>11</sup> It must be mentioned that section three originally pertained to the coding of single variables that speak to risk of bias in observational studies such as those included here. However, because 1) the risk of bias variables had to be constructed based on sparse recommendations just now being brought forth to establish clear standards for study quality assessments in observational research, 2) the lack of validation of the sources and the soundness of their recommended variables, 3) the difficulty in determining the answer to the selected questions in most studies, 4) failure to report the necessary data in several studies, and 5) the lack of any significant variability in the resulting data (e.g., 100% of studies reporting recruiting their participants from the same cohorts), lead to the total exclusion of these variables from all coding and analyses. Appendix C contains all information regarding the issue related to assessing risk of bias in the current studies.

*arrests vs. any returns to custody*), source of recidivism data (e.g., *official records vs. self-report*), measurement of outcome (i.e., *continuous [time-to-recidivism]* vs. *dichotomous* measurement), length of follow up (months at risk), and type of analysis (i.e., effect size *univariate vs. multivariate* in origins). Finally, this section included, for both subgroups, (sub)sample size(s) and labels, variables assessing independence of subsamples and recidivism outcomes, the number (%) of women who recidivated, and finally, the effect size measure (e.g., Odds Ratios (*ORs*) vs. Cohen's *d*), effect size (i.e., the value itself), and direction of the effect (i.e., the meaning of a *positive vs. negative* effect size). The fourth and final portion of the coding sheets is identical to the third section in terms of all initial background factors, as well as the effect size measure, value, and direction. The only difference in coding from Table 3 in section 3 to Table 4 in section 4 is the coding of one overall sample.

### **Coding Decisions: A Brief Note**

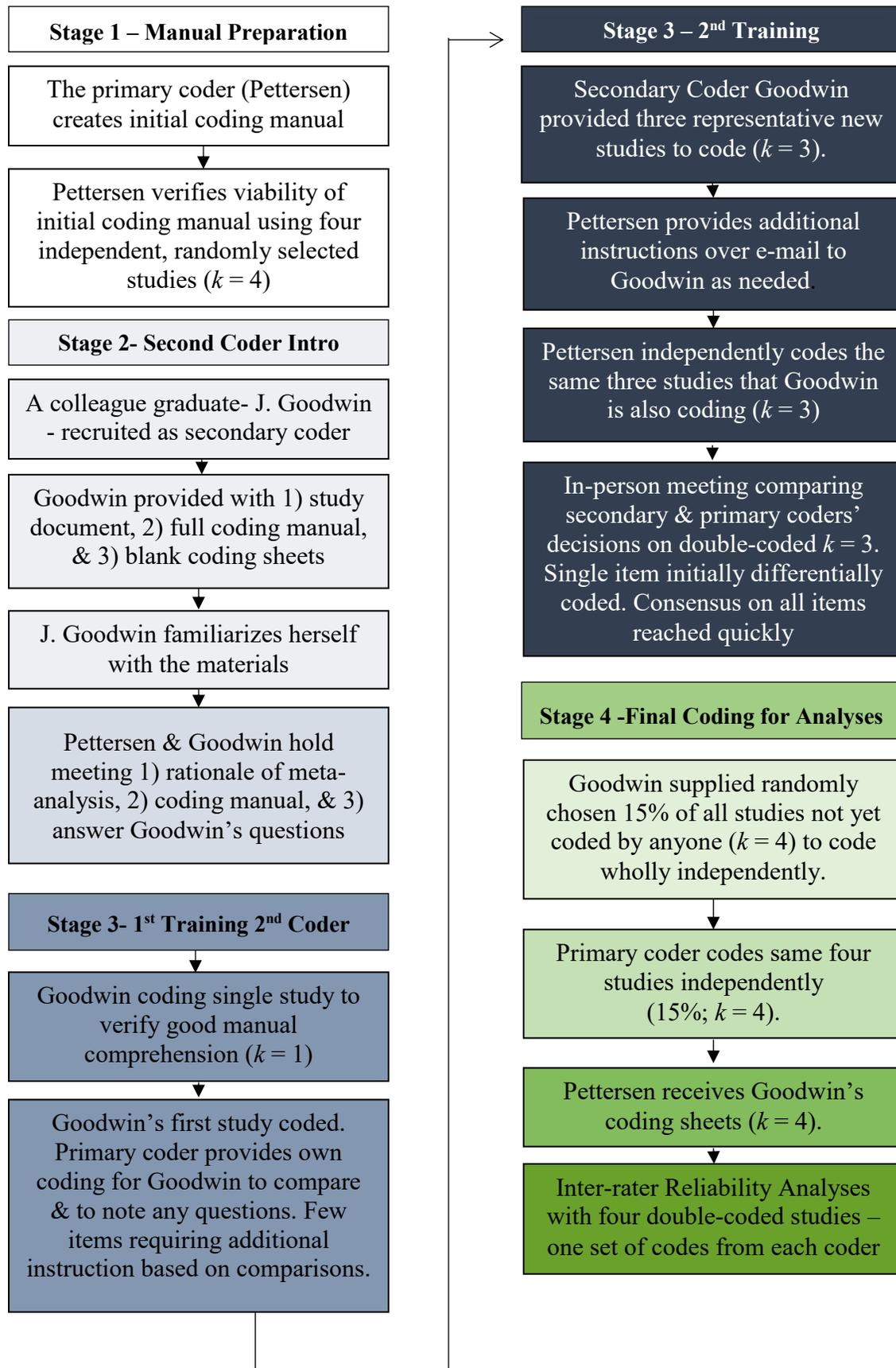
General indicators of mental health (e.g., previous psychiatric treatment) and diagnoses presented in the DSM-IV/DSM-V (APA, 2000; 2013) were coded. The study was not limited to assessment tools specifically adapted for any version of the DSM however, nor to dichotomous measures of mental disorder. If study authors reported continuous scores on a self-report measure of depression, the effect size data pertaining to the association between depression scores and recidivism outcome(s) were also coded. Most importantly, due to the heterogeneous nature of the studies retained for the meta-analysis, and the often large number of effect sizes available from a single study, sample, and measures, a number of individual-study decisions had to be made regarding the selection of an effect size when multiple options were available, as well as what constituted optimal coding of several of the variables across sections the four sections of the manual during the coding process (see Appendix D for detailed coding decisions).

### **Inter-Rater Reliability**

Figure 1 directly below describes the process through five stages: Stage 1) Preparation of the Manual, Stage 2) Introducing the secondary coder, Stage 3) the 1<sup>st</sup> training session with the secondary the coder, Stage 4) the second round of training of the secondary coder, and Stage 5) the final double-coding (i.e., coding by both primary and secondary coder) of 15% ( $k = 4$ ) of all remaining studies to be included in the meta-analysis. The fifteen percent- or four distinct studies- were randomly selected and coded completely independently in a double-blind fashion, were found to represent the variety within the remaining included studies and the literature reviewed as a whole. All studies were eligible for inter-rater analyses. For categorical variables, inter-rater reliability was indexed by Cohen's Kappa, which was calculated using standard procedures in Microsoft Excel (2003). Positive values of Kappa indicate a level of agreement between the coders. Negative values indicate disagreement. Kappa values exceeding .40 are considered indicative of moderate levels of agreement. Inter-rater reliability for continuous variables was assessed using the Interclass Correlation Coefficient (ICC), using a two-way mixed model. The software program SPSS was used for this purpose. ICCs can vary from -1 to 1, with higher values indicating greater inter-rater reliability. An ICC of 0 (zero) indicates agreement at the chance level. Drift- systematic changes in coding guideline application over time (from the first study coded to the last)- was also assessed.

**Figure 1**

*Reliability- Manual, Coder Training, & Inter-rater Reliability Analyses*



This macro was used to compute effect sizes based on both fixed-effects and random-effects models for all mental disorders and indicators of mental health problems separately. It is important to note that in order to aggregate effect sizes, a minimum of three wholly independent effect sizes are required for each predictor (Helmus & Babchishin, 2013). For some mental health indicators, less than three independent effect sizes were available, making the effects for the relevant predictor(s) impossible to aggregate. The following mental health predictors had less than three independent effect sizes available for meta-analyses: psychotropic medication use, compliance with psychotropic medication, suicide attempts as separate from non-lethal self-harm and vice versa, and outpatient psychiatric treatment history. Thus, these variables were not examined in the meta-analysis. Additionally, given that the number of available studies was relatively modest, some mental health indicators were combined and analyzed together. For example, effect sizes for self-harming behaviours and effect sizes for suicidal thoughts and attempts were combined into a single self-harm predictor.

As mentioned above, both fixed-effects and random-effects models were used. Given that the studies included diverse samples, measurement error was not expected to be the sole source of effect size variability. Thus, the random effects model was expected to be the most appropriate measurement model (Borenstein et al., 2009). Random effects models produce more conservative estimates of variances, standard errors, and confidence intervals than do fixed effects models. Random effects models also ensure that the weights assigned to the studies do not cause large studies to dominate the analyses or smaller studies to be trivialized (Borenstein et al., 2009). Although the random-effects model is likely to be superior in the context of the current study, calculating fixed-effects was also necessary, because only this model allows for moderator analyses to be conducted in the statistical software package selected for this study

(i.e., SPSS; Helmus, 2020; Helmus & Babchishin, 2013). In conclusion, both fixed effects and random effects models are reported in the Results section below; when between-study variability exceeds what is expected by measurement error alone, the focus should be on results of random-effects models, but when variability is low, as well as in the context of moderator analyses, the fixed effects model is more appropriate.

The purpose of a meta-analysis is to provide an accessible, succinct quantitative summary of the evidence for a purported effect, association, or parameter, considering all available research findings together. In order to quantitatively describe the relevant effects as well as their variability (i.e., precision or accuracy), a common metric is needed to represent the average effect across studies; this common metric is referred to as an effect size (Borenstein et al., 2009). Unlike null-hypothesis significance testing, an effect size and its associated confidence intervals (*CIs*), allow for the estimation not only of the probability of obtaining the findings if the true effect in the population is zero (0), but also the magnitude, direction, and precision of that estimated effect (Nagakawa & Cuthill, 2007). Of note, for the purposes of the current study, the effect size type or metric (e.g., the Odds Ratio [*OR*]) will be referred to as the effect size *measure*, its value simply as the *effect size*, and the generalized interpretation of the magnitude of the effect as the *relative effect size* (Olivier, May, & Bell, 2917). The paragraph directly below provides details on the selection of a common effect size for the current meta-analysis.

### **Indexing the Average Effect: Choice of Effect Size Measures**

In selecting an effect size measure, the theoretical nature of the latent psychological construct (e.g., *binary vs. continuous*), measurement and scaling of both predictor (e.g., *binary diagnoses of mental disorder vs. continuous self-report scores*) and outcome variables (*binary vs. continuous*), as well as established convention in the area of research (Helmus & Babchishin,

2012a; Polanin & Smilstveit, 2016) are important issues to consider (Coe, 2002). Measurement affects the substantive meaning and interpretability of a resulting aggregate effect size, and some effect size measures' validity is restricted to variables with certain measurement properties. Reporting effect size measures in line with convention ensures that findings can be directly compared across studies and that the majority of the research consumers easily understand their meaning.

In the current study, mental health predictors were often binary, but also not infrequently categorical and continuous (Van Voorhis et al., 2010; 2012; 2013). Recidivism outcomes were always binary, given that the effect sizes derived from continuous time-to-recidivism analyses (Hazard Ratios; *HRs*) cannot be aggregated with other effect sizes from analyses with binary outcomes (Tierney et al., 2007). In terms of convention, the Odds Ratio (*OR*) was the most commonly reported effect size measure in the studies included in analyses in the current meta-analysis. Thus, initially, *ORs* were considered the effect size measure of choice, given its suitability to binary outcomes, any predictor measurement, and both prospective and retrospective designs studies (Hedges & Tipton, 2010). However, an important limitation of the *OR* relevant to the current study, is that it does not accommodate predictors differing in terms of measurement (e.g., combination of binary and continuous predictors). In analyses involving such differently scaled predictors, the resulting magnitude of the *OR* becomes uninterpretable (Nagakawa & Cuthill, 2007; Ranganathan et al., 2017). Briefly, with binary predictors and outcomes, the *OR* is easily understood; and *OR* of 1.5 indicates that odds of recidivism were one-and-a-half times greater for one group relative to the comparison group. *ORs* can vary from 1- the absence of any effect- to infinity (Hedges & Tipton, 2010; Olivier, May, & Bell, 2017). An *OR* below 1 indicates decreased odds while an *OR* above 1 indicates increased odds (Helmus &

Hanson, 2011). With continuous predictors, the *OR* is the change in the outcome associated with a single unit change in the predictor. Thus, the impossibility of interpreting *ORs* derived from a combination of binary and continuous predictors is clear. Despite the limitations of the *OR* in such scenarios, some analyses nevertheless included only predictors with the same measurement scale (i.e., either all binary indicators or all continuous indicators). Thus, given its suitability to the current research context, the *OR* was reported alongside other effect size measures (i.e., Cohen's *d* and the *CLES*) when appropriate (i.e., when the predictors were either all binary or all continuous).

On account of the need for an effect size measure suitable to analyses involving predictor variables sharing measurement scales (e.g., both binary and continuous predictors)- an alternative measure-Cohen's *d*- also known as the standardized mean difference (SMD), was derived alongside the *OR*. Cohen's *d* is believed to be the most popular effect size measure in psychology as a whole, though had been less popular in correctional research wherein odds and risk ratios (*ORs*; *RRs*) are typically preferred. This alternative index is also provided in the main tables displaying meta-analytic results below. Cohen's *d* was selected to represent distinct, individual effects reported in each study included in the aggregate analyses (see Tables presenting the characteristics of the studies involved in each aggregate effect size in the Results section below). Most importantly, while Cohen's *d* is presented for all analyses, the *OR* is additionally included in the main meta-analysis summary table below only when all predictors involved in the analysis were binary. Cohen's *d* is reported in Results tables below whenever differently scaled predictors were involved in analyses. In terms of interpretation, a Cohen's *d* value represents the mean group difference in the outcome in terms of number of standard deviations (*SD*). A Cohen's *d* of 0.5 would indicate that the mean of group one is one half *SD*

larger than the mean of group two (Goulet-Pelletier & Cousineau, 2018). Of note, Cohen (1988; 1992) originally suggested that a  $d = 0.20$  constitutes a small effect, while 0.50 and 0.80 are considered moderate and large effect, respectively.

### **Uses of General Cutoffs: A Cautionary Note**

This utilization of conventional guides to effect size interpretation requires a brief cautionary note. The cutoffs provided by Cohen (1992) are meant to provide some cautious guidance only in the event that other reference points are unavailable. They are not absolute nor meant to be used in isolation from other indices aiding interpretation (Breaugh, 2003; Fritz et al., 2012; Lee, 2016; Nakagawa & Cuthill, 2007; Vaacha-Haase & Thompson, 2004). The magnitude of effect sizes must rather be interpreted with the aim of identifying any potential impact the findings may have on the population of study, clinical practice, correctional practice, and policy (Olivier et al., 2017). In practical terms, effect sizes negligible or very small in magnitude according to Cohen's (1992) generalized cutoffs may have enormous implications in certain contexts. For example, if the outcome is premature death, even a very small effect size can be important (Fritz et al., 2012).

In addition to the *OR* and the Cohen's *d*, the Common Language Effect size (*CLES*), sometimes referred to as the Probability of Superiority (*PS*; Fritz et al., 2012), a universal standardized effect size measure first promoted by McGraw and Wong (1992) was also included. This measure is arguably the most intuitive in terms of interpretation. The *CLES* is the probability of obtaining a difference score greater than zero (0) in the distribution of interest. The interpretation of this effect size is rather intuitive. According to McGraw and Wong (1992), if the effect of interest was the difference in height between men and women, a *CLES* of 0.92 multiplied by 100 would mean that in 92 out of 100 cases (i.e., pairwise comparisons), the man

will be taller than the woman (McGraw & Wong, 1992). This index, ranging originally from 0 to 1, multiplied by 100 to further ease interpretation, is reported alongside the Cohen's  $d$  and  $OR$  values in the main meta-analytic summary tables in the Results section below.

### **Preparing for Aggregation: Derivation of individual ORs, Cohen's $d$ , and CLES effects**

When the desired effect size was not reported directly in the original study, the combination of the well-established and highly recommended *Psychometrica* (Lenhard & Lenhard, 2016) and Wilsons's (2001) *Campbell Collaboration* online effect size calculators and converters, and an Excel sheet (2003) manually and carefully crafted by the current author was most frequently relied upon. The online calculators are versatile tools allowing for the entry of many forms of data- consistently producing precise estimates of most effect sizes.

As mentioned, a Microsoft Excel (2003) sheet was created for the purposes both of maintaining a systematic, comprehensive database of all effect sizes entered directly from individual studies (in the absence of preferred contingency tables or proportions), as well as some effect sizes derived using the online tools described above and to calculate any necessary data points that could not be gleaned from the online calculation tools. Given a number of instances wherein the online conversion tools could not derive the necessary effects, the excel sheet was therefore also used to allow for the necessary step-by-step applications of formulas that eventually allowed for the production of the desired effect size data. Thus, when necessary, calculations of  $ORs$  and the additional effect size measures were relied on the formulae gleaned from a number of independent, reliable sources that together covered all the data scenarios present in my selection of studies (see Table 1 in Appendix E for a list of all relevant formulae and sources. Calculations of  $ORs$ , when online tools were not utilized, also required transformations back and forth to log odds. The complete excel sheet detailing every data point

entered, calculation made, formula employed, and source relied upon, is available from the author upon request. An example of a situation not allowing for the use of online calculator tools or pre-existing excel formulae was when the variance of the *OR* was to be calculated based on correlation coefficients Pearson's *r*, but with the variance of the *r* values themselves missing. In this case, the formula for the calculation of the variance of *r* was identified and employed to derive variance of vast numbers of *r* coefficients in a separate excel sheet, before both values and variances could be transformed into reliable estimates of corresponding *ORs* (see Appendix E, Table 1, for this and all other relevant formulae used to calculate, verify, and convert effect sizes and variances for this meta-analysis).

### **Do not go forth blindly: Verifying Derived Effects**

Despite their favorable reviews and evaluations, the current author felt the need to ensure that the underlying formula upon which the online and calculators' algorithms rested, and naturally, the resulting effect size values and confidence intervals, was indeed sound. Thus, the applicable formula for the calculator-derived *OR* values were identified and the effect size calculated using one-step-at-a-time manual excel calculations. Consistency between calculator outputs and manual effect size calculations was 100%.

### **The Weighted Average Effect size: A Crucial Caveat**

Once effect sizes for individual effects had been derived, aggregation constituted the next step. In meta-analysis, we do not simply calculate an average effect size. Rather, the resulting aggregated effect size is a weighted average. In the current meta-analysis, following established standards (Borenstein et al., 2009), each derived effect size was weighted by the inverse of its variance. Weighting studies by the inverse of their variance is preferable to weighting by sample size alone. Under perfect conditions, when the groups involved have identical sample sizes and

frequencies, weighting by sample size and the inverse of the variance produces equivalent results. However, under sub-optimal conditions- conditions that characterize most if not all studies deemed eligible for inclusion the current meta-analysis, such as incomparable base rates across groups, highly variable subsample sizes, and vast differences in measurement precision. This means that the smaller the variance of an individual effect size, the bigger the weight assigned to it. This way more weight is given to more precise effect size than broad, uncertain estimates (Helmus & Babchishin, 2013).

### **Significance of the Weighted Mean Effect size: Confidence Intervals**

To determine the significance of the weighted mean effect size and to determine its precision, confidence intervals (CI) are utilized. CIs are defined as the range of values in the population with a given probability. Assuming 95% CIs are to be calculated, if the CIs were calculated 100 times, 95% of the resulting CIs would contain the true value in the population (Nakagawa & Cuthill, 2007). Favorable CIs are those that, if many were to be calculated, would contain the true effect size in the population (Goulet-Pelletier & Cousineau, 2018). Put differently, of a total of 100, 95% CIs, 95 would include the true population mean (Lee, 2016). CIs are therefore presented alongside Cohen's  $d$  effect sizes.

### **Study Heterogeneity and Outliers**

To assess between-study heterogeneity and the presence of outliers, I used a combination of the Cochran's  $Q$  and  $I^2$  statistics. The  $Q$  statistic is distributed as a chi-square with  $k-1$  degrees of freedom (Hedges & Olkin, 1985). If measurement error is the only source of between-study heterogeneity,  $Q$  can be expected not to exceed its associated degrees of freedom. It is worth noting that the magnitude of  $Q$  is influenced by the number of included studies. Thus, the value of  $Q$  cannot meaningfully be compared across analyses that contain different numbers of studies.

Instead, to test whether  $Q$  is significant, meaning that between-study variability is unlikely to be due to measurement error alone, the value of  $Q$  is compared against the critical value associated with the desired significance level (i.e.,  $p = .05$  in the current context). In short,  $Q$  can inform us of the statistical significance of between-study variability, but not the magnitude of that variability.

The  $I^2$  is the percentage of error above chance. Unlike  $Q$ ,  $I^2$  is not dependent on the number of included studies and can be compared across analyses. It can vary between 0 and 100. Higgins, Thompson, Deeks, and Altman (2003) suggested that 25%, 50%, and 75% can be considered *low*, *moderate*, and *high* degrees of variability, respectively. Importantly, some studies may produce extreme effect sizes that have a meaningful impact on the resulting overall aggregate effect sizes, potentially distorting the overall results of the meta-analysis. Although there are no universally agreed-upon conventions outlining best practices for managing outliers in meta-analyses, Hanson and Bussière (1998) suggested that if 1) the outlier is the most extreme value (i.e., the effect size is either the largest or the smallest among the effect sizes included in the aggregate effect size), 2) the overall  $Q$  is significant, and 3) it accounts for more than 50% of the total variability in the aggregate effect, removing the offending effect size may be warranted. Helmus and Babchishin (2013) elaborated on these rules, stating that an outlying effect size, in addition to having the largest or smallest value of those included in the specific analysis, it should also have the most extreme weighted squared deviation. In the current study, in accordance with Helmus and Babchishin's (2013) practices, all analyses were conducted with and without the outlier(s) included and results of both analyses were reported.

## Moderators

A number of the variables coded were originally considered for moderator-analyses. Moderation is present when the relationship between a predictor and an outcome depends on the level of the moderating variable. For example, if the relationship between BPD and recidivism depends on the risk of the individual woman, then risk would be considered a moderator. It is important to note that in order to examine a moderator, at least three independent effect sizes must be available for each level of the moderator (Helmus & Babchishin, 2013).

A-priori planned moderator analyses included: 1) publication status (*yes vs. no*), 2) peer-review (*yes vs. no*), 3) study design (*retrospective vs. prospective*), and 4) study location, the following sample characteristics 1) predominant disorder (i.e., the most common diagnosis of mental disorder in the sample), 2) average age, 3) average or majority risk level, 4) sample index offense (i.e., *majority violent vs. non-violent index offense[s]*), 5) majority race (*white vs. non-white*), 6) assessment setting (*federal/state prison vs. local jail vs. community*), 7) measurement of the mental health predictor (*dichotomous vs. continuous*), 8) *current vs. lifetime* diagnoses, and 9) assessment method (all methods compared).

In reality, only a small number of moderators could be examined in the current study. There were several reasons for this. First, a relatively limited number of studies and effect sizes were available for meta-analyses in the current study, resulting in moderator variables that did not have a sufficient number of independent effects for each level of the moderator (categorical moderators); See the Results section below for the moderators that had to be excluded on these grounds. Second, moderator analyses require significant variability: conducting moderator analysis requires, at minimum, a moderate  $I^2$  value, and some methodologists argue that both a significant  $Q$  and a moderate to large  $I^2$  value is required (Helmus & Babchishin, 2011). In the

current meta-analysis, a moderate  $I^2$  was considered adequate to justify moderator analyses. In some analyses, this criterion was not met, and moderator analyses were not conducted.

Relatedly, some potential continuous moderators showed a less than ideal distribution (see Results section below for the moderators that were excluded on these grounds). Categorical moderators such as assessment method and publication status were examined in an analysis of variance- like procedure using the macro from Helmus and Babchishin (2013), producing a between-level  $Q$ . A between-level  $Q$  is the amount of overall variability that is explained by the moderator (Helmus & Babchishin, 2013). The total variability is divided into a portion that can be explained by the moderator, and a residual portion, expressed as  $Q_{between}$  and  $Q_{within}$  values. A significant  $Q_{between}$  value indicates that the difference between categories is larger than what can be expected by sampling error alone (Lipsey & Wilson, 2001).

For continuous moderators (e.g., year of study completion), a regression-like procedure presented in Helmus and Babchishin (2013) was used. This involved running a linear regression with the effect size as the dependent variable, the derived weight variable as the weighted least squares (WLS), and the moderator as the independent variable. Note that while the resulting unstandardized regression coefficient ( $b$ ) in the resulting SPSS output is correct, but the associated standard error ( $SE$ ) is too big, requiring a calculation to obtain its correct value (see formula directly below). The corrected  $t$ -test using the corrected  $SE$  was then calculated (see formula below).

$$\text{Corrected } SE = \frac{SE_{b1}}{\sqrt{MSE}}$$

*Formula for calculation of the corrected SE (Helmus & Babchishin, 2013).*

$$\text{Corrected } t - \text{test} = \frac{b}{SE_{\text{Corrected}}}$$

*Formula for calculation of the corrected t-test (Field, 2009).*

### **Independence of Observations**

Independence of observations is a very important concept in meta-analyses (Helmus & Babchishin, 2012; Pratt & Cullen, 2000). Effect sizes are non-independent if the relevant (sub)group(s) included in analyses overlap partially or completely with another (sub)sample for which an effect size is available, whether the same or different outcome measures were used. For example, a study may report an effect size for ADHD and *general recidivism* followed by an effect size for ADHD with *violent recidivism* specifically. In this case, violent recidivism is not a distinct, independent outcome but rather actually represents a subset of crimes under the general recidivism umbrella term. Thus, in this case, the effect sizes resulting from the analyses are non-independent. Only one of the effect sizes can be included in a single aggregate effect size. Importantly, this does not mean that effect sizes from partially or completely overlapping samples cannot be used when separate analyses are conducted for different mental health predictors. Using the example above, if one analysis examined the relationship between ADHD and general recidivism and another analysis examined the relationship between BPD and general recidivism, and provided there were at least three independent effect sizes for each of these predictors, two separate meta-analyses resulting in two separate aggregate effect sizes would be conducted for ADHD and BPD predictors. Thus, in each analysis, only one set of proportions and/or one effect size was included in each analysis.

Importantly, a number of studies reported a large number of effect sizes pertaining to identical predictors and outcomes. When this occurred, the included effect size was selected to maximize sample size, length of follow-up, and the comprehensiveness of the recidivism outcome. Effect size selection strategy and criteria followed by detailed accounts of each and

every study-specific effect size selected and included in all utilized predictor variables are available in Appendix F.

### **Preliminary Findings: Literature Search and Inter-rater Reliability**

This Results section is organized to reflect the order of presentation of the independent section in the sections above. First, accounts of manual searches not described in the Methods section below, along with results of author inquiries, a note regarding discrepancies between current and typical reporting strategies (i.e., a note on overlapping records), and lastly, a summary of the literature search followed by three figures presenting a streamlined graphical view of the process are provided. Next, the results of inter-rater reliability analyses are discussed in some detail. A summary of all reliability tests is provided in Table 2 in the Inter-rater Reliability Analysis section below. Subsequently, the section detailing the major meta-analytic findings begins, again with an overview of the organizational structure of the report.

### **Literature Search Summary**

The end date of the literature search was March 31<sup>st</sup>, 2020. The sections below briefly outline results of manual searches and of personal communication with primary study authors. In addition, a brief note describing a discrepancy between current reporting and typical reporting of meta-analytic literature searches, specifically in terms of overlapping records, is provided.

Finally, a summary of findings and three supporting figures are provided.

### **Manual searches**

In addition to the main literature search strategies described in detail in the Methods Section above, manual searches were also conducted. Reference lists of a substantial number of previous meta-analyses, systematic reviews, as well as all primary studies eligible for inclusion in the current study, were scoured for additional records. One reference list pertaining to the

Bonta et al. (2014) meta-analysis discussed in depth in the Introduction section above, requires brief mention here. All 73 studies in this reference list were tracked down and retained for further inspection. Despite the relevance of this prior meta-analysis to the current study, this additional search was largely unsuccessful in identifying additional eligible records. All but three studies were rejected; the vast majority failed to aggregate by gender and some had subsamples of women smaller than the minimum 10. The three initially retained studies all utilized MDO samples. These were initially considered eligible and retained for coding. However, upon careful reflection on the appropriateness of the comparison groups utilized, two of three were subsequently excluded (see the post-hoc exclusion criteria in the Method section above for a detailed account of the reason for this exclusion).

### **Author Inquiries**

When key data were missing or additional information was needed from already included primary studies, the original study authors were contacted. Overall, either the primary investigator or the corresponding author of 47 distinct studies were contacted via e-mail. When delivery attempts were unsuccessful, one additional attempt was made using the other author's e-mail instead. However, for successfully delivered messages, only one attempt was made per study. The authors were provided three months to respond before the inquiry was recorded as closed. Roughly 50% of all authors contacted responded. Inquiries included attempts to obtain unpublished univariate data, gender-aggregated data, or additional contextual information (e.g., base rates, method of measurement, recidivism definition). In addition, a few inquiries were made simply in efforts to obtain authors' clarification regarding the correct interpretation of reported findings. A single Norwegian author was contacted to obtain a document pre-publication, but this was unsuccessful as analyses had not yet been completed (i.e., Dr. Bukten and colleagues'

extensive study of mental health and addictions among incarcerated offenders in Norway). Taken together, the inquiries resulted in one additional unpublished document providing complementary data to an already included study (i.e., Carroll [2017] was provided to support the Odio et al. [2018] publication), and three sets of previously unpublished univariate data pertaining to three independent studies. Finally, a single inquiry resulted in the rejection of a study initially considered eligible; in this study the temporal order of measurement or predictor and outcome was violated; Janssen et al., 2017). Appendix G provides a list of authors and their contact information as well as results of inquiries in tabular form.

### **A Note on Overlapping Records**

Some notable obstacles occurring during the literature search required initial search strategy and terms to be adjusted. First, in terms of the academic databases search, the search engine did not allow for the simultaneous entry of an the very large number of search terms. Thus, the overall search was broken up into multiple searches involving clusters of fewer, related terms. It is therefore possible that even within one source (e.g., the academic database) some overlapping records were screened more than once. Noting the title and the abstract of every record that was identified across a large number of separate searches in order to verify how many studies were identified more than once, did not constitute a defensible strategy in terms of time management or indeed in terms of the value of the potentially obtained information. Further, collecting the titles and abstracts occurring for all clusters of search terms, across all sources, including many sources lacking a search engine of any sort, some sites requiring manual “scrolling only” searches, and some failing to not provide any end point to the search, was not deemed to be a viable option. Indeed, in combination with the enormous number of records identified and screened across sources (i.e., > 700 in academic databases, > 8, 000 in other

sources), these obstacles prohibited systematic, precise, and comprehensive identification of the absolute number of totally or partially overlapping records. The unavoidable failure to arrive at an accurate count of overlapping studies is relevant to the interpretation of the graphical figures presented below. Note that when overlap was identified and the information contained in the two distinct documents was deemed complimentary (e.g., the earlier unpublished thesis associated with an already collected published study contained important additional data excluded from the published manuscript), both records were kept, but only the initially identified document was counted as a unique record.

### **Summary Report**

The searches of the academic databases resulted in the identification of 751 records, some of which may have overlapped. Based on titles and abstracts, all but 26 were excluded. Screening the 26 full-texts resulted in the rejection of 14 more. Reasons for exclusions included problematic designs (e.g., temporal order of measurement was reversed; cross-sectional design;  $k = 4$ ), lack of data for women ( $k = 3$ ), lack of a recidivism outcome ( $k = 2$ ), absence of mental health predictors ( $k = 2$ ), inappropriate non-criminal justice sample ( $k = 1$ ), inappropriate youth sample ( $k = 1$ ), and inadequate data to derive an effect size ( $k = 1$ ). Twelve studies remained eligible for inclusion, but of these, four were rejected at the coding stage: unreliable reporting ( $k = 1$ ), all recidivists sample ( $k = 1$ ), and only multivariate findings ( $k = 2$ ). Thus, the exhaustive academic database search resulted in eight studies suitable for inclusion in the quantitative synthesis. Figure 2 below is a graphic representation of the academic database search from start to finish. The searches of the alternative sources, taken together, resulted in the initial identification of 8, 751 records. Recall that a uniqueness of records cannot be guaranteed. Alternative sources included alternative databases, academic, government, and professional

networking websites, the CBMH academic journal, Google Scholar, and manual searches. Based on inspection of titles and abstracts, 114 were initially retained for full text screening; 97 of which were subsequently rejected and excluded from the meta-analysis<sup>12</sup>.

Following this screening process, 17 studies remained eligible for inclusion. Of these, another seven were rejected at the coding stage; five of these studies used inappropriate comparison samples and two reported only multivariate findings. In total, ten studies were found suitable for inclusion in the current quantitative synthesis. Figure 3 below is a graphic representation of the alternative source search from start to finish.

Additionally, Figure 4 is presented directly following Figures 2 and 3, serving as an easy reference point and simple summary displaying the results of the entire literature search and records evaluation process across all sources. In total, 18 independent studies to be included in the current meta-analysis. A number of these studies were associated with partially overlapping documents that contained supplementary data. Naturally, these documents were not counted towards the reported total number of unique contributions.

Of note, the number of studies rejected based on careful review of full-text documents was 140 and with the addition of associated studies, brief reports, annexes, supplements, and appendices the total number rose to 150. Appendix H contains a comprehensive reference list of full-texts reviewed and excluded, including those studies rejected after initial inclusion during

---

<sup>12</sup> The 97 rejected records were excluded on account of problematic designs (e.g., temporal order of measurement was reversed; cross-sectional design;  $k = 2$ ), lack of data for women ( $k = 62$ ), the sample of women was smaller than the minimum 10 required ( $k = 5$ ), lack of a recidivism outcome ( $k = 8$ ), absence of mental health predictors ( $k = 19$ ), inadequate comparison groups ( $k = 5$ ), inappropriate non-criminal justice sample ( $k = 5$ ), and inadequate data to calculate an effect size ( $k = 8$ ).

the formal coding process (i.e., Erickson, 2016; Hines-Randolph, 2014), along with a brief note following each individual reference outlining the reason(s) for its rejection.

### **Results of Inter-rater Reliability Analyses**

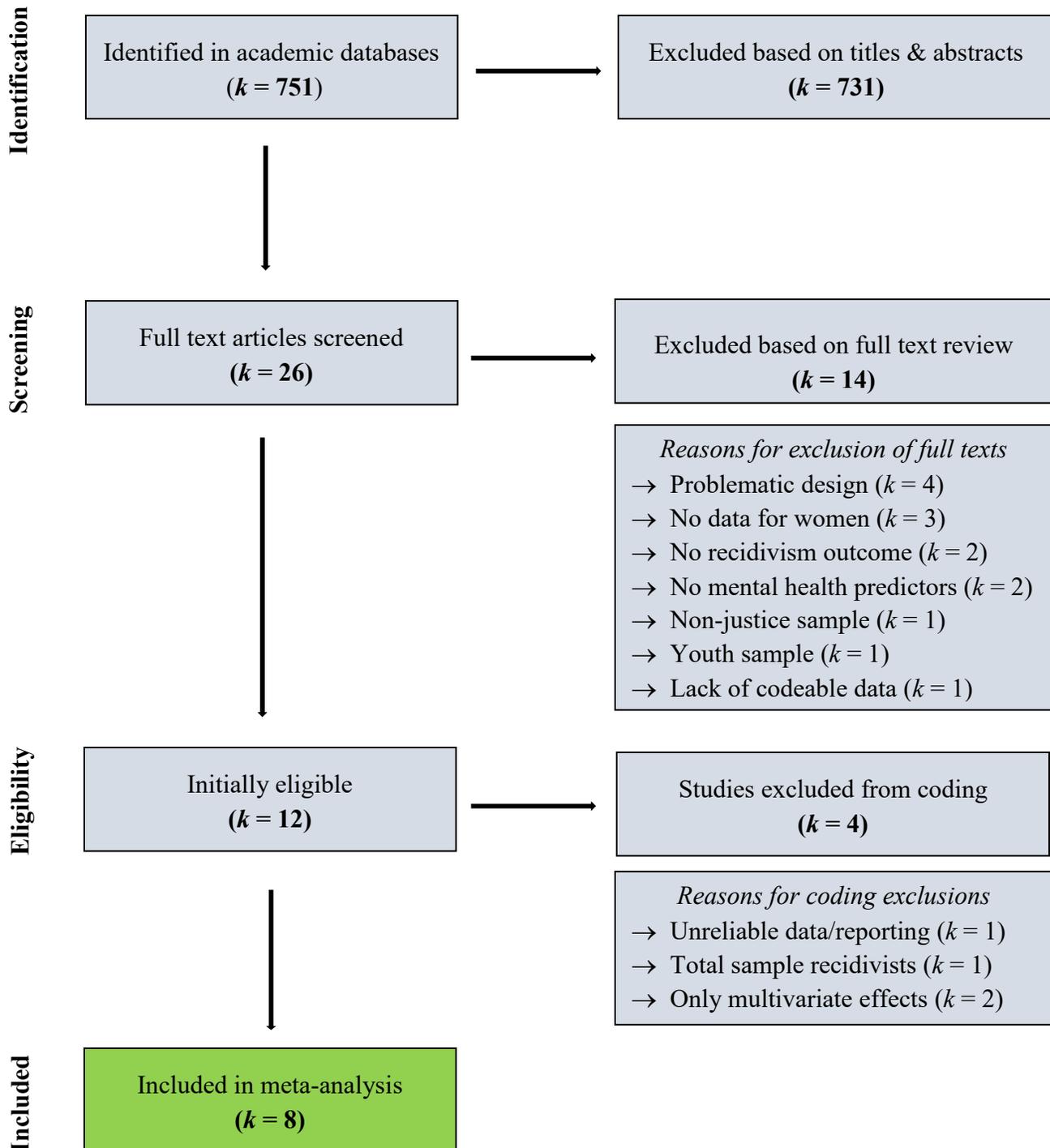
As explained in detail in the Methods section above, Kappa and ICC indices were calculated for categorical and continuous variables, respectively. Four (approximately 15%) of the studies were coded by both coders and were used to examine inter-rater reliability. Appendix I provides a list of variables excluded from analyses as well as strategies employed to avoid biasing analyses (i.e., artificially inflating apparent agreement between coders).

In addition, data on three variables were only reported in a single study. In these cases, inter-rater reliability was examined manually (i.e., visually). Tables 2 below present ICC values, Kappa values, and degree of agreement for study and sample characteristics and for effect size data (Table 3) for all relevant variables. In addition, the variables for which assessment was purely visual are marked in each table along with levels of agreement obtained for these special cases. Table 2 below presents results for sample and study characteristics. Perfect inter-rater agreement was achieved for ninety percent of all sample- and study-related variables coded (18/20). In two instances, moderate and moderate-high agreement was achieved.

The setting in which the mental health assessment was conducted (e.g., community, local jail, state/federal prison, or forensic mental health unit/hospital) achieved moderate to high agreement, as indicated by a Kappa of 0.64. Inconsistent coding occurred for a single study; one coder identified both community and local/remand jail (mixed recruitment) as the setting(s), while the other identified only the community. Disagreement was immediately resolved upon review of the study. For the three remaining studies, agreement was perfect.

**Figure 2**

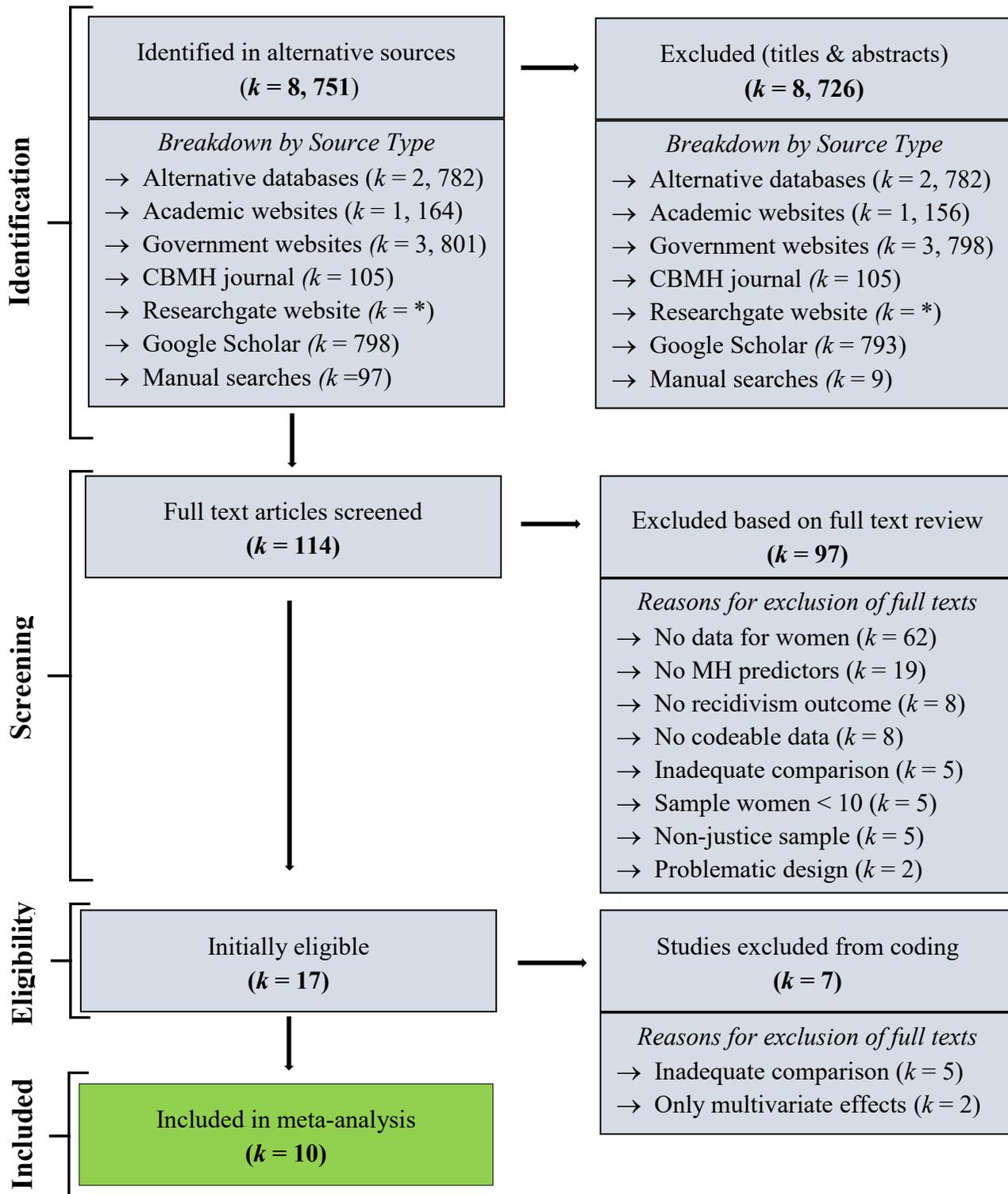
*Literature Review Flow Chart: Academic Databases*



*Note.* As explained in detail above, the search engines, number of terms, and multitude of sources did not permit for derivation of the number of overlapping and/or unique studies.

**Figure 3**

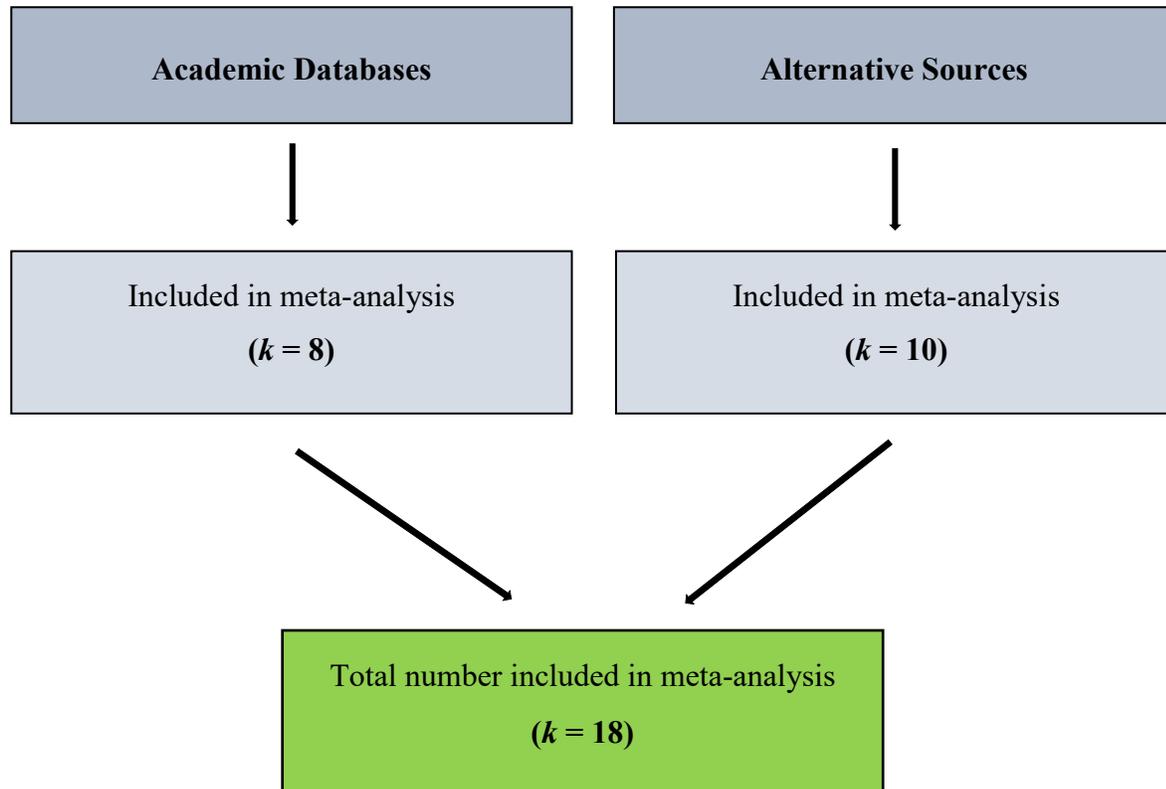
*Literature Review Flow Chart: Alternative Sources*



*Note.* Alternative = non-English sources. Identification of unique records was not possible. \* Website does not provide number of records. MH = Mental health.

**Figure 4**

*Literature Review Flow Chart: All Included Studies from all Sources*



*Note.* As explained in detail above, the search engines, number of terms, and multitude of sources did not permit for derivation of the number of overlapping and/or unique studies.

Risk assessment type was the only other study and sample-related variable for which perfect agreement was not obtained. The coding conflicted as the risk assessment type was identified as Actuarial in one case and Structured Professional Judgement (SPJ) in the other. Adequate information about the tool was not provided in the primary study. One coder made the

correct judgment based on pre-existing knowledge. The correct answer was verified after a brief literature search. Because disagreement on this variable only occurred in a single study, a moderate overall level of agreement was still obtained.

Inter-rater reliability for effect size data are presented in Table 3. Perfect agreement was achieved for 93.33% of these variables (14/15). Low inter-rater agreement was obtained for the mental health assessment method variable. Agreement was perfect on this variable in three of four studies, indicating that coders generally understood how to assess the variable and follow the coding protocol guidelines consistently. Nevertheless, in a single study, eight effect sizes were reported, and one coder's misidentification of assessment method resulted in complete disagreement for all eight effects, having an unexpectedly strong negative effect on the overall agreement rating. One coder identified self-report as the sole assessment method, while the other identified the combination of both self-report and file review as the method. The error was corrected and agreement immediately achieved upon review of the study.

In summary, the inter-rater reliability analyses supported the coding protocol as a reliable tool for data extraction allowing for consistency over time and between coders. In cases of disagreement, coders reached complete consensus after immediately following analyses. No signs of meaningful, systematic drift over time in coding guideline applications could be identified; inter-rater agreement did not appear to depend on whether the studies were coded at outset, mid-way through the coding process, or towards the end. Thus, the current author concluded that the coding manual functioned as intended and that the data entered into the dataset for use in meta-analyses were reliable.

**Table 3**

*Inter-Rater Agreement (Reliability): Study & Sample Characteristics (Kappa & ICC)*

<b>Study characteristics</b>	<b>Kappa</b>	<b>ICC</b>	<b>Agreement</b>
Study design (retrospective vs. prospective)	1	-	perfect
Study location (country)	1	-	perfect
Peer-reviewed (yes vs. no)	1	-	perfect
Recidivism base rate(s)	-	1	perfect
<i>Setting (of mental health assessments)</i>	<i>0.636</i>	-	<i>moderate-high</i>
<b>Sample characteristics</b>			
Total N (before attrition)	-	1	perfect
Mean age (before attrition)	-	1	perfect
Predominant mental disorder (e.g., depression)	1	-	perfect
Percentage (%) diagnosed (all mental disorders) <sup>a</sup>	-	-	perfect
Risk type assessed (e.g., violent)	1	-	perfect
<i>Risk assessment type (e.g., actuarial)</i>	<i>0.429</i>	-	<i>moderate</i>
Name of Risk Assessment <sup>a</sup>	-	-	perfect
Average risk level	1	-	perfect
Majority risk level	1	-	perfect
Index offenses (% non-violent, violent, sexual) <sup>b</sup>	-	-	-
Majority race (white, non-white, mixed, or unknown)	1	-	perfect
White (%)	-	1	perfect
Black (%)	-	0.998	perfect
Hispanic/Latina (%)	-	1	perfect
Asian/Pacific Islander <sup>b</sup>	-	-	-
Indigenous (%)	-	1	perfect
Other race (%)	-	1	perfect

*Note.* Variables for which the level of inter-rater reliability was meaningfully lower than 1 (100%) are presented in *italics*. “-” = value/index not relevant to the variable. 1 = perfect agreement. <sup>a</sup> Data for variable was only available from a single study and agreement was determined by manual visual inspection. <sup>b</sup> One study was selected to code the variable, resulting in

**Table 4**

*Inter-Rater Agreement (Reliability): Effect Size Data (Kappa & ICC)*

<b>Effect Size Data</b>	<b>Kappa</b>	<b>ICC</b>	<b>Agreement</b>
Mental health predictor	1	-	perfect
Mental health predictor scale (continuous vs. binary)	1	-	perfect
<i>Mental health assessment method</i>	<i>0.265</i>	<i>-</i>	<i>low</i>
Name of mental health assessment protocol <sup>a</sup>	-	-	perfect
Timing of diagnoses (current vs. lifetime)	1	-	perfect
Recidivism type (e.g., general vs. violent)	1	-	perfect
Recidivism definition (e.g., breaches vs. returns)	1	-	perfect
Recidivism source (i.e., official vs. unofficial)	1	-	perfect
Recidivism measurement (continuous vs. binary)	1	-	perfect
Length of follow-up	-	1	perfect
Sample size ( <i>n</i> ) in analyses	-	1	perfect
Sub-group (non)independence	1	-	perfect
Number of recidivists <sup>b</sup>	-	-	-
Number of non-recidivists <sup>b</sup>	-	-	-
Type of effect size	1	-	perfect
Direction of effect (positive vs. negative)	1	-	perfect
Effect size value	-	1	perfect

*Note.* Variables for which the level of inter-rater reliability was meaningfully lower than 1 (100%) are presented in *italics*. “-” = value/index not relevant to the variable. 1 = perfect agreement. Kappa and ICC values below moderate in value are presented in **red font**. <sup>a</sup> Data for variable was only available from a single study and agreement was determined by manual, visual inspection. <sup>b</sup> One coder neglected to code the variable, excluding it from analyses.

### Main Results: What the Quantitative Synthesis Tells Us

The current section details the results of the meta-analysis proper. First, a brief presentation of mental health variables that were included in the current meta-analysis is required<sup>13</sup>. As the Literature Search Results section above indicates, the number of empirical, primary studies that met current inclusion criteria was limited. As a natural consequences of sparse data, the mental health predictors examined in the current analyses became restricted. A minimum of three independent effect sizes had to be available for analyses to be possible. The following predictors were retained for analyses: 1) depression ( $k = 11$ ), 2) anxiety ( $k = 11$ ), 3) PTSD ( $k = 4$ ), 4) psychosis ( $k = 7$ ), 5) personality disorder (PD;  $k = 4$ ), 6) any mental disorder (with and without severe/major disorders;  $k = 6$  and  $k = 4$ ), 7) self-harm ( $k = 4$ ), and 8) psychiatric history ( $k = 3$ )<sup>14</sup>.

The first predictor of interest was depression. The following section introduces findings pertaining to this predictor, alongside two summaries (Tables 4 and 5) providing study and sample characteristics, effect size data, and individual Cohen's  $d$  (95% CI) effect sizes. Finally, Table 6 presents the aggregate effect sizes in random- and fixed- effects models, as well as indices of between-study variability ( $Q$ -between and  $I^2$ ). Effects are presented for all depression indicators (with and without outliers), effects for binary diagnoses, and effects limited to depression scores.

---

<sup>13</sup> The current study excluded APD, Psychopathy, and Substance Use/Dependence Disorders.

<sup>14</sup> Predictors of initial interest also Schizophrenia ( $k = 0$ ), ADHD (any subtype;  $k = 1$ ), Bipolar disorder(s) ( $k = 1$ ), Dissociative Disorders ( $k = 0$ ), Adjustment Disorders ( $k = 0$ ), Narcissistic Personality disorder (NPD;  $k = 0$ ), BPD ( $k = 2$ ), and Self-harm and Suicide separately, but none of these met the minimum number of effect sizes necessary to allow for analyses. Of note, two effect sizes for BPD were included in aggregate estimated effects for personality disorders in general (PD).

### Depression and General Recidivism

There were 11 independent effect sizes available for depression. Depression included diagnosis of major depressive disorder (MDD;  $k = 2$ ; Study ID# = 1, 6), dysthymia (DY;  $k = 1$ ; Study ID# = 3), depressive disorder not otherwise specified (DD-NOS;  $k = 4$ ; Study ID# = 9, 21, 14, 5), scores on a measure of depression ( $k = 3$ ; Study ID# = 4, 22, 23), and scores on a measure of risk of developing depression ( $k = 1$ ; Study ID# = 17).

Table 5 below presents the most basic study and sample characteristics while Table 6 presents recidivism information, sample sizes, and attrition rates. It also presents the Cohen's  $d$  effect sizes and their associated 95% confidence intervals for each individual study. The effect sizes represent the relationships between depression and general recidivism.

Table 7 below displays the average Cohen's  $d$  effect size and associated 95% confidence intervals, Odds Ratios ( $ORs$ ), and  $CLES$  effect sizes for all predictors of general recidivism examined in this study with the current predictor of interest remaining depression. Measures of between-study variability ( $Q$ -between and  $I^2$ ) are also reported; Significant values indicate variability beyond measurement error alone. Importantly, this table reports effect sizes for the main categorical moderator, which was predictor measurement (i.e., binary vs. continuous). In addition, effects reported in this table are presented with and without any identified outlier. Results of all additional categorical moderator analyses are reported in Table 8 below.

**Table 5**

*Depression: Study and Sample Characteristics (k = 11)*

Study [ID#]	Design	Peer-review	Country	Setting	Sample	Race (%)	Mean Age
Blanchette (1996)[#1] <sup>a</sup>	P	Yes	CA	CF	-	White (74%)	32.9
Ducat et al. (2017)[#9] <sup>b</sup>	P	Yes	AU	CF	FS	-	34.6
Erickson (2014; 2016)[#21]	R	No	U.S.	CF	LR	Mix (32%)	35.8
Light et al (2013)[#17] <sup>c</sup>	P	No	U.K.	CF	-	White (83%)	31.2
Mannerfelt & Håkansson (2018)[#5]	R	Yes	SE	CF	SA	White (-)	36.2
Papadopulos (2011)[#3]	R	No	U.S.	CF	-	White (78%)	-
Scott et al. (2014; 2016) [#4]	P	Yes	U.S.	CF	SA	Black (8%)	-
Tripodi et al. (2019) [#6]	P	Yes	U.S.	CF	-	Mix (5%)	33.7
Van Voorhis et al.(2007-10)[#14] <sup>d</sup>	P	Yes	U.S.	CF/COM	-	Mix (30%)	34.3
Van Voorhis et al. (2012)[#22]	P	No	U.S.	CF/COM	-	White (71%)	33.9
Van Voorhis et al. (2013)[#23]	P	No	U.S.	C	-	White (64%)	33.9

*Note.* “-” = no special sample features or missing data. P = Prospective; R = Retrospective. Peer Review = Minimum one document per study was peer-reviewed. CA = Canada; AU = Australia; U.S. = United States; U.K. = United Kingdom; SE = Sweden. CF = Correctional facility. COM = Community. FS = All fire-setters; LR = all low risk; SA = all substance abusers; Race = majority (≥ 60%); Mix = No majority race. <sup>a</sup> Included overlapping document (Blanchette & Motiuk, 1996[#1]). <sup>b</sup> Recidivism = fire-related offenses only. <sup>c</sup> Included overlapping document (Ministry of Justice, 2013-Annex A [#17]). <sup>d</sup> Included four partially overlapping studies/documents (Van Voorhis et al., 2007; 2008; 2009; 2010 [#14]).

**Table 6**

*Depression: Recidivism, Attrition, Sample Sizes, & Cohen’s d Effect Sizes (95% CI)*

Study [ID#]	Recidivism	Follow-up	Base rate	% attrition	n	Cohen’s d (95% CI)
Blanchette (1996) [#1] <sup>a</sup>	RCUS	24	.59	13	66	0.28 (95% CI; -0.21, 0.77)
Ducat et al. (2017) [#9] <sup>b</sup>	NC	82	.07	24	143	0.58 (95% CI; -0.15, 1.32)
Erickson (2014; 2016) [#21]	REA	24	.26	0	294	0.08 (95% CI; -0.15, 0.30)
Light et al. (2013) [#17] <sup>c</sup>	REC	12.	.54	0	116	<b>0.81 (95% CI; 0.37, 1.26)*</b>
Mannerfelt & Håkansson (2018) [#5]	RCJS	x	.62	0	407	-0.02 (95% CI; -0.24, 0.21)
Papadopulos (2011) [#3]	RCUS	36	.28	0	150	0.24 (95% CI; -0.08, 0.56)
Scott et al. (2014; 2016) [#4]	RCUS	12	.32	0	387	-0.01 (95% CI; -0.27, 0.26)
Tripodi et al. (2019) [#6]	RCUS	36	.18	0	230	0.25 (95% CI; -0.01, 0.51)
Van Voorhis et al. (2007-10) [#14] <sup>d</sup>	REA	24	.22	0	158	0.18 (95% CI; -0.13, 0.49)
Van Voorhis et al. (2012) [#22]	REA	12	.27	4	601	<b>0.20 (95% CI; 0.04, 0.36)*</b>
Van Voorhis et al. (2013) [#23]	AF	12	.39	6	553	0.10 (95% CI; -0.07, 0.27)

*Note.* **Bolded** effects =  $p < .05$ . RCUS = return to custody; NC = New charge; REA = Re-arrest; REC = Re-conviction; RCJS = Return to criminal justice system; FREC = Felony re-conviction; AF = Any fail (any REA, REC, RCUS, and any technical breaches). Follow-up = months. n = sample in analyses.

**Table 7**

*All Mental Health Predictors & General Recidivism: Aggregate Effect Sizes & Between Study Variability*

Mental Health Predictor	<i>k</i>	<i>N</i>	Cohens <i>d</i> (95% CI)		<i>OR</i>	<i>CLES</i> * 100	<i>Q</i> ( <i>Q</i> -between)	<i>I</i> <sup>2</sup>
			Fixed Effects	Random Effects				
Depression	11	3,105	<b>0.15 (0.08, 0.23)*</b>	<b>0.17 (0.07, 0.27)*</b>	x	54	15.41	35.10
Without outlier	10	2,989	<b>0.13 (0.06, 0.21)*</b>	<b>0.13 (0.06, 0.21)*</b>	1.27	53	6.83	0.00
Binary diagnoses	7	1,448	<b>0.14 (0.03, 0.25)*</b>	<b>0.14 (0.03, 0.25)*</b>	x	54	4.91 (0.08)	0.00
Continuous scores	4	1,657	0.16 (0.06, 0.27)	0.21 (-0.01, 0.42)	x	54	10.42* (0.08)	71.22*
Anxiety	11	3,284	0.08 (0.01, 0.14)	0.10 (-0.01, 0.20)	x	53	19.72*	49.29*
Binary diagnoses	7	1,337	0.09 (-0.03, 0.21)	0.07 (-0.11, 0.26)	1.18	53	11.49 (0.05)	47.77
Continuous scores	4	1,947	0.07 (0.00, 0.15)	0.11 (-0.03, 0.26)	x	53	8.18* (0.05)	63.34*
Excluding PTSD	7	1,877	0.02 (-0.06, 0.10)	0.04 (-0.07, 0.15)	x	52	7.92	24.23
PTSD Only (All)	4	1,407	<b>0.18 (0.08, 0.29)*</b>	<b>0.18 (0.01, 0.35)*</b>	x	55	5.70	47.41
Psychosis	7	1,825	<b>0.14 (0.04, 0.25)*</b>	0.14 (-0.02, 0.29)	x	54	9.91	39.46
Personality disorder	4	1,210	0.14 (0.03, 0.25)	<b>0.37 (0.02, 0.71)*</b>	x	58	18.19*	83.51*
Without outlier	3	1,067	0.11 (0.00, 0.22)	0.20 (-0.05, 0.44)	x	53	7.35*	72.79*
Any Mental Disorder	6	2,964	<b>0.31 (0.14, 0.49)*</b>	<b>0.33 (0.12, 0.53)*</b>	1.79	58	6.11	18.15
Self-harm & Suicidality	4	4,246	<b>0.08 (0.02, 0.14)*</b>	0.05 (-0.06, 0.15)	x	52	5.94	49.52
Psychiatric History	3	667	<b>0.33 (0.10, 0.56)*</b>	<b>0.36 (0.04, 0.68)*</b>	x	59	2.95	32.30

*Note.* The primary model is the random effects model; for significant effects, bold font indicates that the effect size is interpretable, while standard font indicates that *d* from the fixed effects model is disregarded due to excessive between-study variability, regardless of significance. When *d* is not significant, bold font is not used. PTSD = post-traumatic stress disorder.

\* Significant at *p* < .05.

Continuous moderator analyses are presented last, and only in text. The resulting main aggregate effect size representing all 11 depression predictors was significant but very modest in magnitude in both models. Contrary to hypotheses suggesting that binary diagnoses would more strongly indicative of mental dysfunction relative to continuous scores, results revealed that the effect size did not depend on predictor measurement. A single exception occurred; when analyses were limited to continuous scores it fell just barely below significance in the random effects model. A single outlier was identified (Light et al. 2013; Study ID#17) and when it was excluded, between-study variability was adequately reduced, meaning remaining variance may be accounted for by measurement error alone. Taken together, the lack of meaningful impact on effect size significance, magnitude, or the accuracy of estimation (i.e., confidence intervals remained narrow), further supported depression as a very modest but significant predictor of recidivism.

Results of all supplementary categorical moderator analyses are included in Table 7 above, provided an adequate number of studies and effect sizes were available. Notably, due to inadequate independent effect sizes for the remaining predictors, only depression and anxiety had are addressed in this secondary summary table. Importantly, the current focus on depression notwithstanding, this table also includes results for all other predictors included in the current study. Detailed accounts of findings for the remaining predictors are presented in systematic fashion following Table 7. Only race was found to be a significant moderator, with results suggesting that depression significantly predicted recidivism only for samples in which the majority of women were White (> 60%). Data allowed for examination of length of follow-up and publication year as continuous moderators; neither were significant ( $t(df=8) = 1.67, p > .05, ns$ ;  $t(df=8) = 0.89, p > .05, ns$ , respectively).

**Table 8**

*All Supplementary Categorical Moderator Analyses: Aggregated Fixed Cohen's d Effects: All Predictors & General Recidivism*

Predictor and Moderator Category	<i>k</i>	<i>N</i>	<i>Cohens d</i> (95% CI)	<i>CL</i> * 100	<i>Q</i>	<i>(Q-between)</i>	<i>I</i> <sup>2</sup>
Depression (Majority White)	5	1, 600	0.18 (0.08, 0.28)*	53	1.52	(4.52)*	0.00
Depression (Majority Non-White)	3	839	0.08 (-0.08, 0.23)	51	0.79		0.00
Depression (Lifetime Assessment)	4	766	0.12 (-0.05, 0.29)	54	3.85	(0.04)	22.12
Depression (Current Assessment)	6	2, 223	0.14 (0.05, 0.22)*	52	2.94		0.00
Depression (Peer-reviewed)	6	1, 391	0.11 (-0.01, 0.24)	53	5.28	(0.17)	5.28
Depression (Non-Peer-reviewed)	4	1, 598	0.15 (0.05, 0.25)*	53	1.38		0.00
Depression (Retrospective Design)	3	851	0.07 (-0.07, 0.22)	52	1.62	(0.98)	0.00
Depression (Prospective Design)	7	2, 138	0.16 (0.07, 0.25)*	54	4.23		0.00
Anxiety (Majority White)	5	1, 542	0.18 (0.08, 0.28)*	52	7.85	(7.90)*	49.03*
Anxiety (Majority Non-White)	4	1, 192	0.01 (-0.08, 0.10)	52	3.97		24.45
Anxiety (Lifetime Assessment)	5	813	0.05 (-0.12, 0.21)	53	10.54*	(0.12)	62.04*
Anxiety (Current Assessment)	6	2, 471	0.08 (0.01, 0.15)*	53	9.06		44.81
Anxiety (Peer-reviewed)	6	1, 514	-0.03 (-0.11, 0.06)	51	4.52	(12.23)*	0.00
Anxiety (Non-Peer reviewed)	5	1, 770	0.20 (0.11, 0.30)*	54	2.97		0.00
Anxiety (Retrospective Design)	3	851	0.15 (0.00, 0.30)	53	4.79	(1.25)	58.26
Anxiety (Prospective Design)	8	2, 433	0.06 (-0.01, 0.13)	52	13.68		48.83

*Note.* Only depression and anxiety had an adequate number of independent effect sizes for each level of the moderator and therefore the only predictors presented in this table. Depression analyses excluded the outlying effect size. \* Significant at *p* < .05.

**Anxiety and General Recidivism**

There were 11 independent effect sizes available for anxiety. Anxiety predictors included all diagnoses of anxiety disorders generalized anxiety disorder (GAD;  $k = 1$ ; Study ID#1), Post-traumatic stress disorder (PTSD;  $k = 2$ ; Study ID#3 and 13), anxiety disorders NOS ( $k = 4$ ; Study ID#5, 6, 9, and 21), continuous scores on general measures of anxiety ( $k = 1$ ; Study ID#17), anxiety and fear symptoms ( $k = 1$ ; Study ID#4), and PTSD ( $k = 2$ ; Study ID#22 and 23).

Table 9 presents basic study and sample characteristics. Table 10 presents  $d$  for each study and other contextual effect size data (e.g., base rates).

**Table 9**

*Anxiety ( $k = 11$ ) and PTSD ( $k = 4$ ): Study and Sample Characteristics*

Study [ID#]	Design	Peer-review	Country	Setting	Sample	Race (% White)	Mean Age
Blanchette & Motiuk (1996)[#1] <sup>a</sup>	P	Yes	CA	CF	-	White (74%)	32.9
Ducat et al (2017)[#9] <sup>b</sup>	P	Yes	AU	CF/COM	FS	-	34.6
Erickson (2014; 2016)[#21]	R	No	U.S.	CF	LR	Mix (32%)	35.8
Kubiak (2004)[#13] <sup>d</sup>	P	Yes	U.S.	CF	SA	Mix (50%)	38.1
Light et al (2013)[#17] <sup>c</sup>	P	No	U.K.	CF	-	White (83%)	31.2
Mannerfelt & Håkansson (2018)[#5]	R	Yes	SE	CF	SA	White (x)	36.2
Papadopulos (2011)[#3] <sup>d</sup>	R	No	U.S.	CF	-	White (78%)	-
Scott et al. (2014; 2016)[#4]	P	Yes	U.S.	CF	SA	Black (8%)	-
Tripodi et al. (2019)[#6]	P	Yes	U.S.	CF	-	Mix (54%)	33.7
Van Voorhis et al. (2012)[#22] <sup>d</sup>	P	No	U.S.	CF	-	White (71%)	33.9
Van Voorhis et al. (2013)[#23] <sup>d</sup>	P	No	U.S.	COM	-	White (64%)	34.2

*Note.* “ - “ = no special sample features or missing data. P = Prospective; R = Retrospective. Peer Review = Minimum one document per study was peer-reviewed. Country: CA = Canada; AU = Australia; SE = Sweden. Setting: CF = Correctional facility. COM = Community. FS = all fire-setters; LR = all low risk; SA = all substance abusers; Race = majority ( $\geq 60\%$ ); Mix = No majority race. <sup>a</sup> Included overlapping study Blanchette & Motiuk (1996)[#1]. <sup>b</sup> Recidivism = only fire-related offenses (e.g., arson). <sup>c</sup> Included overlapping supplementary document Ministry of Justice (2013)-Annex A [#17]. <sup>d</sup> Predictor = PTSD.

**Table 10**

*Anxiety, PTSD, & General Recidivism: Cohen’s d Effect Sizes, Outcome, & Sample Size Data (k = 11)*

Study [ID#]	Recidivism	Follow-up	Base rate	% Attrition	n	Cohen’s d (95% CI)
Blanchette & Motiuk (1996)[#1] <sup>a</sup>	RCUS	24	13	13	66	0.28 (95% CI; -0.21, 0.77)
Ducat et al (2017)[#9] <sup>b</sup>	NC	82	24	24	143	0.58 (95% CI; -0.15, 1.32)
Erickson (2014; 2016)[#21]	REA	24	0	0	294	0.08 (95% CI; -0.15, 0.30)
Kubiak (2004)[#13] <sup>d</sup>	REC	X	22	0	47	<b>0.81 (95% CI; 0.37, 1.26)*</b>
Light et al (2013)[#17] <sup>c</sup>	RCJS	12	0	0	116	-0.02 (95% CI; -0.24, 0.21)
Mannerfelt & Håkansson (2018)[#5]	RCUS	X	0	0	407	0.24 (95% CI; -0.08, 0.56)
Papadopulos (2011)[#3] <sup>d</sup>	FREC	36	0	0	150	-0.01 (95% CI; -0.27, 0.26)
Scott et al. (2014; 2016)[#4]	RCUS	36	0	0	621	0.25 (95% CI; -0.01, 0.51)
Tripodi et al. (2019)[#6]	RCUS	36	0	0	230	0.18 (95% CI; -0.13, 0.49)
Van Voorhis et al. (2012)[#22] <sup>de</sup>	AF	6	0.002	4	625	<b>0.20 (95% CI; 0.04, 0.36)*</b>
Van Voorhis et al. (2013)[#23] <sup>de</sup>	AF	6	0	6	585	0.10 (95% CI; -0.07, 0.27)

*Note.* RCUS = return to custody; NC = New charge; REA = Re-arrest; REC = Re-conviction; RCJS = Return to criminal justice system; FREC = Felony re-conviction; AF = Any fail (any REA, REC, RCUS, and any technical breaches). Follow-up = months. n = sample in analysis. **Bolded** effect sizes are significant at  $p < .05$ . \* Significant at  $p < .05$ .

Table 7 presented above displays the aggregate Cohen’s d, OR, and CLES effect sizes for all predictors of general recidivism including the current overlapping predictors of interest; anxiety and PTSD and results of analyses of the main categorical moderator of interest- predictor measurement; predictor measurement. Recall that results of all supplementary categorical moderator analyses are reported in the second major summary Table 8 above. Lastly, continuous moderator analyses are in text.

As observed in the primary summary Table 7 above, anxiety was not significantly associated with general recidivism in the main analysis. Q indicated significant between-study variability, but no outliers were identified. Excluding PTSD from analyses did not affect the resulting effect size. The effect did not depend on how anxiety was measured; binary and continuous measurement yielded comparable effect sizes. In terms of supplementary categorical moderator analyses presented in Table 8 above, results suggested that anxiety may significantly predict recidivism only in samples consisting mainly of White women (> 60%). No other

significant moderation effects were identified. Nevertheless, the fact that small, significant effects were uncovered when analyses were restricted to effects involving only current mental health assessments and when studies were not peer reviewed, contradicting results of main analyses ( $d = 0.08$  to  $d = 0.20$ ). Data allowed for examination of length of follow-up and publication year as continuous moderators; neither were significant ( $t(df=8) = 1.67, p > .05, ns$ ;  $t(df=8) = 0.89, p > .05, ns$ , respectively). In summary, anxiety appeared unrelated to recidivism across all main analyses (refer to Table 7 above), but moderator results suggested that this relationship may be dependent on other factors, including study methodology. It is worth noting however, that the initial hypothesis predicting that anxiety would emerge as a potential protective factor was consistently *not* supported.

### **Post-traumatic stress disorder (PTSD) and General Recidivism**

When PTSD was excluded from the main anxiety analysis. Instead, this disorder was examined in isolation. The reasons for its separation from anxiety as a whole are as followed: 1) the etiology of typical anxiety disorders such as GAD and PTSD are believed to be different – PTSD is caused by exposure to severe trauma, 2) the symptoms of PTSD relative to other anxiety disorders are so profoundly different from other anxiety disorders that it would be difficult to justify hypotheses suggesting their behavioural and life outcomes would be the comparable, 3) risk/needs correctional assessments measures and addresses PTSD as a condition wholly distinct from other anxiety disorders, and 4) due to the tremendous differences between cause, expression of, experience of, and reported outcomes of anxiety disorders in general relative to PTSD, it is possible that their potential relationships to criminal behaviour may be juxtaposed (i.e., if combined into a single predictor, null findings, or completely unreliable conclusions may possibly occur). Interestingly however, the initial hypothesis pertaining to

PTSD specifically was supported; effect sizes were significant and indicated that PTSD may be associated with increased recidivism (refer to main summary Table 7 above).

**Psychosis and General Recidivism**

There were seven independent effect sizes available for psychosis. Psychosis included psychotic disorders NOS ( $k = 3$ ; Study ID#6, 9, and 14) and symptoms associated with psychoses ( $k = 4$ ; Study ID#5, 17, 22, and 23). Table 11 presents basic study and sample characteristics. Table 12 presents  $d$  for each study and other contextual effect size data (e.g., base rates).

**Table 11**

*Psychosis: Study and Sample Characteristics (k = 7)*

Study [ID#]	Design	Peer-review	Country	Setting	Sample	Race (% White)	Mean Age
Ducat et al (2017)[#9]	P	Yes	AU	CF	FS	-	34.6
Light et al (2013)[#17] <sup>a</sup>	P	No	U.K.	CF	-	White (83%)	31.2
Mannerfelt & Håkansson (2018)[#5]	R	Yes	SE	CF	SA	-	36.2
Tripodi et al. (2019)[#6]	P	Yes	U.S.	CF	-	Mix (54%)	33.7
Van Voorhis et al. (2007-10)[#14] <sup>b</sup>	P	Yes	U.S.	CF	-	White (80%)	33.8
Van Voorhis et al. (2012)[#22]	P	No	U.S.	CF	-	White (71%)	33.9
Van Voorhis et al. (2013)[#23]	P	No	U.S.	COM	-	White (64%)	34.2

*Note.* “-” = no special sample features or missing data. P = Prospective; R = Retrospective. Peer Review = Minimum one document per study was peer-reviewed. AU = Australia; U.K. = United Kingdom; SE = Sweden. U.S. = United States. CF = Correctional facility. COM = Community. FS = all fire-setters; LR = all low risk; SA = all substance abusers; Race = majority ( $\geq 60\%$ ); Mix = No majority race. <sup>a</sup> Included overlapping supplementary document (Ministry of Justice, 2013-Annex A [#17]). <sup>b</sup> Included four overlapping studies (Van Voorhis et al., 2007, 2008, 2009, 2010).

Table 7 above displays the aggregate effect sizes for psychosis as well as for all other predictors of general recidivism. The aggregate effect size for psychosis fell below threshold for statistical significance in the random effects model. Data did not allow for any moderator analyses.

**Table 12**

*Psychosis & General Recidivism: Individual Cohen’s d Effect Sizes, Outcome, & Sample Size Data (k = 7)*

Study [ID#]	Recidivism	Follow-up	Base rate	% attrition	n	Cohen’s d (95% CI)
Ducat et al (2017)[#9] <sup>a</sup>	NC	82	.07	24	143	0.61 (95% CI; -0.31, 1.54)
Light et al (2013)[#17] <sup>b</sup>	REC	12	.54	0	115	0.40 (95% CI; -0.08, 0.87)
Mannerfelt & Håkansson (2018)[#5]	RCJS	X	.62	0	407	-0.01 (95% CI; -0.31, 0.29)
Tripodi et al. (2019)[#6]	REC	36	.18	0	230	0.12 (95% CI; -0.14, 0.38)
Van Voorhis et al. (2007-10)[#10] <sup>c</sup>	REC	24	x	10	244	-0.14 (95% CI; -0.39, 0.11)
Van Voorhis et al. (2012)[#22]	OF	12	.37	4	601	<b>0.24 (95% CI; 0.08, 0.40)*</b>
Van Voorhis et al. (2013)[#23]	AF	12	.78	9	83	0.30 (95% CI; -0.13, 0.74)

*Note.* NC = New charge; REC = Re-conviction; RCJS = Return to Criminal Justice System; OF = Offense fail (any re-arrest, re-conviction, or technical violations). AF = Any fail (any re-arrest, re-conviction, re-incarceration, or technical breaches). Follow-up = months. n = sample in analysis. <sup>a</sup> Recidivism = fire-related offenses (e.g., arson). **Bolded** effect sizes are significant at  $p < .05$ . \* Significant at  $p < .05$ .

**Personality disorder (PD) and General Recidivism**

There were only four independent effect sizes available for personality disorder (PD). PD included PD NOS ( $k = 2$ ; Study ID#3 and 9) and BPD symptoms ( $k = 2$ ; Study ID#4 and 21). Table 13 presents basic study and sample characteristics. Table 14 presents  $d$  for each study and other contextual effect size data (e.g., base rates).

Treating personality disorder (PD) as a unitary construct may not be optimal, given that authors in the two studies examining PD NOS did not describe the nature of these disorders: Diagnoses may or may not relate meaningfully to criminal conduct. Nevertheless, the decision was made to make full use of the sparse data available and conduct exploratory analyses. Table 13 below demonstrates that three out of four individual effects were significant and positive, supporting further exploration.

Table 7 presented above displays the aggregate Cohen’s  $d$ ,  $OR$ , and  $CLES$  effect sizes PD- the current predictor of interest-as well as for all other predictors of general recidivism. Given that between-study variability was excessive, the focus here remains on the random effects model. In

this model, a significant effect emerged; the magnitude of this effect fell in the small to moderate range ( $d = 0.37$ ).

**Table 13**

*PD: Study and Sample Characteristics (k = 4)*

Study [ID#]	Design	Peer-review	Country	Setting	Sample	Race (% White)	Mean Age
Ducat et al (2017)[#9]	P	Yes	AU	CF	FS	-	34.6
Erickson (2014; 2016)[#21]	R	No	U.S.	CF	LR	Mix (32%)	35.8
Papadoplous[#3]	R	No	U.S.	CF	-	White (78%)	-
Scott et al. (2014; 2016)[#4]	P	Yes	U.S.	CF	SA	Black (9%)	36.7

*Note.* “-” = no special sample features or missing data. P = Prospective; R = Retrospective. Peer Review = Minimum one document per study was peer-reviewed. AU = Australia; U.S. = United States. CF = Correctional facility. FS = all fire-setters; LR = all low risk; SA = all substance abusers; Race = majority (≥

Study [ID#]	Recidivism	Follow-up	Base rate	% attrition	<i>n</i>	Cohen’s <i>d</i> (95% CI)
Ducat et al (2017)[#9] <sup>a</sup>	NC	82	.07	24	143	<b>1.46 (95% CI; 0.66, 2.25)*</b>
Erickson (2014; 2016)[#21]	REA	24	.26	0	294	<b>0.27 (95% CI; 0.04, 0.50)*</b>
Papadoplous[#3]	REC	36	.28	0	150	<b>0.41 (95% CI; 0.08, 0.74)*</b>
Scott et al. (2014; 2016)[#4]	REC	36	.70	0	623	0.01 (95% CI; -0.13, 0.14)

60%); Mix = No majority race.

**Table 14**

*PD & General Recidivism: Cohen’s d Effect Sizes, Outcomes, & Sample Sizes (k = 4)*

*Note.* NC = New charge; REA = Re-arrest; REC = Re-conviction. Follow-up = months. *n* = sample in analysis.

<sup>a</sup> Recidivism = fire-related offenses (e.g., arson). **Bolded** effect sizes are significant at  $p < .05$ . \* Significant at  $p < .05$ .

A clear outlier was detected. Given the magnitude of the difference between this individual effect and the effects of the remaining three studies (refer back to Table 13 directly above) and the incomparability of the weighted least squares (WLS) deviations (i.e., the outlier had a WLS deviation of 10.63 while the remaining studies ranged from 1.25 to 3.60), the

analyses were re-run excluding the outlier. Rather than resulting in a larger positive aggregate effect and narrower confidence intervals however, a smaller ( $d = 0.20$ ) effect resulted, which fell below statistical significance. While between-study variability remained excessive, I did not attempt to remove subsequent outliers given that only three studies remained (Helmus & Babchishin, 2013). Naturally, no moderators could be examined with the limited number of effect sizes available. In conclusion, the data available for inclusion were so limited, and the outcome of the analyses were inconsistent at best, no firm conclusion can be drawn regarding the effect of PD on general recidivism.

**Any Mental Disorder and General Recidivism**

There were six independent effect sizes available for the *any mental disorder* as a predictor of general recidivism. Predictors included diagnosis of any major mental disorder (MMD;  $k = 2$ ; Study ID# = 1 and 10)<sup>15</sup>, diagnosis of an axis I disorder ( $k = 1$ ; Study ID# = 9)<sup>16</sup>, any history of mental disorder (MD;  $k = 1$ ; Study ID# = 11), and simply as any mental disorder (undefined;  $k = 2$ ; Study ID# 7 and 14). Table 15 presents basic study and sample characteristics.

Study [ID#]	Design	Peer-review	Country	Setting	Sample	Race (% White)	Mean Age
Blanchette (1996)[#1] <sup>a</sup>	P	Yes	CA	CF	-	White (74%)	32.9
Ducat et al (2017)[#9]	P	Yes	AU	CF	FS	-	34.6
Gehring (2018)[#11]	P	Yes	U.S.	COM	PRT	Mix (49%)	27.1
King et al. (2018)[#10]	P	Yes	U.S.	CF	-	Mix (44%)	42.8
Visher & Bakken (2014)[#7]	P	Yes	U.S.	CF	-	Black (x)	36.1
Van Voorhis et al. (2007-10)[#14]	P	Yes	U.S.	CF	-	Mix (30%)	34.3

Table 16 presents  $d$  for each study and other contextual effect size data (e.g., base rates).

**Table 15**

*Any Mental Disorder (MD): Basic Study & Sample Characteristics*

<sup>15</sup> Study 1 (Blanchette, 1996) defined MMD as Schizophrenia, Schizophreniform Disorder, or Mania. Study 10 (King et al., 2018) defined it as Bipolar disorders, Dissociative Disorder, or Schizoaffective disorder.

<sup>16</sup> Included axis I disorders were Cognitive Disorders, BPD, Depressive Disorder NOS, Bipolar Disorders, Anxiety disorders NOS, Disorders first Diagnosed in Childhood or Adolescence (DMS-IV; APA, 2000), Psychotic disorders NOS, and PD NOS (Ducat et al., 2017).

*Note.* “no special sample features or missing data. P = Prospective study design. Peer Review = Minimum one document per study was peer-reviewed. CA = Canada; AU = Australia; U.S. = United States. CF = Correctional Facility; COM = Community. FS = all fire-setters; PRT = all pre-trial detainees; Race = majority ( $\geq 60\%$ ); Mix = No majority race. <sup>a</sup>Included overlapping study (Blanchette & Motiuk, 1996[#1]).

**Table 16**

*Any Mental Disorder (MD) & General Recidivism: Cohen’s d Effect Sizes, Outcomes & Sample Size Data (k = 6)*

Study [ID#]	Recidivism	Follow-up	Base rate	% attrition	n	Cohen’s d (95% CI)
Blanchette (1996)[#1] <sup>a</sup>	RCUS	24	.59	13	66	0.81 (95% CI; -0.08, 1.69)
Ducat et al (2017)[#9] <sup>b</sup>	NC	82	.07	24	143	<b>1.44 (95% CI; 0.28, 2.59)*</b>
Gehring (2018)[#11]	REA	6	.06	0	103	0.35 (95% CI; -0.05, 0.74)
King et al. (2018)[#10]	RCUS	96	.44	0	2, 311	0.10 (95% CI; -0.33, 0.53)
Visher & Bakken (2014)[#7]	SREA	9	.32	46	77	0.41 (95% CI; -0.002, 0.81)
Van Voorhis et al. (2007-10)[#14]	REA	24	.22	0	158	0.26 (95% CI; -0.05, 0.58)

*Note.* x = missing data. RCUS = Return to custody; NC = New charge; REA = re-arrests; SREA = Self-reported re-arrest. Follow-up is presented in months and has been rounded to a single integer. % attrition has been rounded to one whole integer. n = sample size in analysis. \* **Bolded** Cohen’s d effect sizes were significant at  $p < .05$ . <sup>a</sup> Includes overlapping study (Blanchette & Motiuk, 1996[#1]). <sup>b</sup> General recidivism was restricted it to fire-related offenses (e.g., arson) only.

Table 7 presented above displays the aggregate Cohen’s d, OR, and CLES effect sizes PD-the current predictor of interest-as well as for all other predictors of general recidivism. The aggregate Cohen’s d effect size was significant and small to moderate in size ( $d = 0.33$ ;  $OR = 1.79$ ). Between-study variability did not exceed what could be attributed to measurement error.

The limitations of the aggregate Cohen’s d derived for this predictor must be clarified here. First, included effects were limited to any mental disorder predictors as defined a-priori by primary study authors. The current author did not combine proportions or effects reported for individual disorders. This could not be attempted, as membership in one disorder group was not an exclusion criterion for membership in another disorder group (i.e., the same individual could suffer both PTSD and Depression and contribute to effect size data for both) in the original reports. Thus, if these these effect sizes or proportions had been combined it would have result in non-independence. Second, the definitions of any mental disorder were sometimes missing and differed between remaining studies, making it impossible to attribute the observed effect to any specific diagnostic category

**Self-harm and Suicidality and General Recidivism**

Self-harm was indexed by only four independent effect sizes and included suicidal ideation ( $k = 1$ ), suicide attempts ( $k = 1$ ), and history of self-harm ( $k = 1$ ), and suicidal or homicidal ideation ( $k = 1$ ). Table 17 presents basic study and sample characteristics. Table 18 presents  $d$  for each study and other contextual effect size data (e.g., base rates).

**Table 17**

*Self-harm and Suicidality: Basic Study & Sample Characteristics ( $k = 4$ )*

Study [ID#]	Design	Peer-review	Country	Setting	Sample	Race (% White)	Mean Age
Erickson (2014; 2016)[#21]	R	No	U.S.	CF	LR	White (24%)	35.8
Mannerfelt & Håkansson (2018)[#5]	R	Yes	SE	CF	SA	-	36.2
Olson et al. (2016)[#2]	P	Yes	U.S.	CF	-	Mix (41%)	35.7
Scott et al. (2014; 2016)[#4]	P	Yes	U.S.	CF	SA	Black (9%)	36.7

*Note.* “-“ = no special features or missing data. P = Prospective study design; R = Retrospective study design. Peer-review = minimum one document per study was peer reviewed. SE = Sweden; U.S. = United States. CF = Correctional facility. FS = all fire setters (e.g., arson); SA = all substance abusers. Race = sample majority (60%+); Mix = no majority race.

**Table 18**

*Self-harm, Suicidality, & General Recidivism: Individual Cohen’s  $d$  Effect Sizes, Outcomes, & Sample Sizes ( $k = 4$ )*

Study [ID#]	Recidivism	Follow-up	Base rate	% attrition	$n$	Cohen’s $d$ (95% CI)
Erickson (2014; 2016)[#21]	REA	24	.26	0	294	0.12 (95% CI; -0.11, 0.35)
Mannerfelt & Håkansson (2018)[#5]	RCJS	x	.62	0	407	-0.16 (95% CI; -0.39, 0.07)
Olson et al. (2016)[#2]	REA	38	.61	0	2, 923	<b>0.11 (95% CI; 0.04, 0.19)*</b>
Scott et al. (2014; 2016)[#4]	RCUS	36	.70	0	622	0.01 (95% CI; -0.13, 0.15)

*Note.* REA = re-arrests; RCJS = Return to criminal justice system; REA = Re-arrest; RCUS = return to custody. Follow-up= months.  $n$  = sample size involved in analysis. \* **Bolded** Cohen’s  $d$  effect sizes were significant at  $p < .05$

Table 7 presented above displays the aggregate Cohen’s *d*, *OR*, and *CLES* effect sizes PD-the current predictor of interest-as well as for all other predictors of general recidivism. The aggregate Cohen’s *d* effect size was not significant in the random effects model. Between-study variability was not excessive; a very small significant effect emerged in the fixed effects model.

This finding contradicted a-priori the hypothesis, as well as existing theoretical accounts and empirical findings linking self-directed aggression to both violence and recidivism risk (Power et al., 2013). However, analyses may be underpowered because of the small number of individual effects involved.

**Psychiatric History and General Recidivism**

Psychiatric history was indexed by only three independent effect sizes. Predictors included variables such as *registered with any public mental health service while at risk in the community*. Table 19 presents basic study and sample characteristics. Table 20 presents *d* for each study and other contextual effect size data (e.g., base rates).

**Table 19**

*Psychiatric History: Basic Study & Sample Characteristics (k = 3)*

Study [ID#]	Design	Peer-review	Country	Setting	Sample	Race (% White)	Mean Age
Ducat et al (2017)[#9]	P	Yes	AU	CF	FS	-	34.6
Light et al (2013)[#17] <sup>a</sup>	P	No	U.K.	CF	-	White (83%)	31.2
Mannerfelt & Håkansson (2018)[#5]	R	Yes	SE	CF	SA	-	36.2

*Note.* “-“ = no special features or missing data. P = Prospective study design; R = Retrospective study design. Peer-review = minimum one document was peer-reviewed per study. AU = Australia; U.K. = United Kingdom; SE = Sweden. CF = Correctional facility. FS = all fire setters (e.g., arson); SA = all substance abusers. Race = sample majority (60%+); Mix = no majority race.

**Table 20**

*Psychiatric History & General Recidivism: Individual Cohen’s d Effect Sizes, Outcomes & Sample Sizes*

(*k* = 3)

Study [ID#]	Recidivism	Follow-up	Base rate	% attrition	<i>n</i>	Cohen’s <i>d</i> (95% CI)
Ducat et al (2017)[#9] <sup>a</sup>	NC	82	.07	24	143	<b>1.32 (95% CI; 0.17, 2.47)*</b>
Light et al (2013)[#17] <sup>b</sup>	REC	12	.54	0	117	0.27 (95% CI; -0.13, 0.68)
Mannerfelt & Håkansson (2018)[#5]	RCJS	x	.62	0	407	<b>0.30 (95% CI; 0.004, 0.59)*</b>

*Note.* NC = New charge; REC = Re-conviction; RCJS = Return to the criminal justice system. Follow-up = months. *n* = sample size involved in analysis. \* **Bolded** Cohen’s *d* effect sizes were significant at *p* < .05.

Table 7 presented above displays the aggregate Cohen’s *d*, *OR*, and *CLES* effect sizes Psychiatric history-the current predictor of interest-as well as for all other predictors of general recidivism. The aggregate Cohen’s *d* effect size was significant in both models and small to moderate in size; it was the largest effect size obtained in the current meta-analysis (*d* = 0.36). Between-study variability was not excessive. In summary, it appears that psychiatric history may be a significant predictor of general recidivism. It is crucial to note however, that no firm conclusion can be drawn based on the extremely limited data available in this study (*k* = 3; *N* = 667).

## Discussion

The main goal of the current study was to examine the potential relationship between mental health and recidivism among justice-involved women through quantitative syntheses of the existing literature. The focus of the study was broad, such that both diagnoses of specific mental disorders as well as general indicators of mental health status (e.g., self-harming behaviours) could be addressed.

The literature review and study identification process revealed that empirical evidence speaking to the association between mental health and recidivism among adult women is extremely sparse. Despite the identification of more than 9,000 records, the total number of studies included in the meta-analysis was 18. Nevertheless, the main findings of the meta-analyses are provided here. Results for the following mental health conditions are presented in order below; depression, anxiety, PTSD, psychosis, personality disorder, and any mental health disorder, along with moderator analyses where applicable. Lastly, results for two general indicators; self-harm and suicidality and psychiatric history are presented. Of note, the number of reported effect sizes for violent recidivism was too small to allow for analyses with this outcome. All results below therefore pertain to general recidivism. Additionally, data allowed for examination of moderators only for depression and anxiety.

First, results of analyses supported the hypothesized association between depression and increased recidivism. The effect was modest, however. Contrary to expectations, the relationship did not depend on whether current or lifetime diagnoses were assessed, nor on whether predictors were binary diagnoses or scores on continuous measures of depression were used. Interestingly, race emerged as a significant moderator, suggesting that depression as measured in

the included studies only predicts recidivism in samples wherein the majority of women were White (>60%).

Second, contrary to the hypothesis that anxiety would emerge as a protective factor, results did not support any association between anxiety and recidivism. Two significant moderators emerged for anxiety: race and peer-review status. When analyses were limited to a) samples in which the majority of women were White and b) effects reported in non-peer reviewed studies, anxiety was found to be significantly associated with increased recidivism. Third, as opposed to results derived when anxiety was examined as a unitary construct (i.e., when all anxiety disorders and measures were included), analyses limited to PTSD supported the initial hypothesis; PTSD was modestly but positively related to recidivism. Fourth, psychosis was not supported as a predictor of general recidivism. Unfortunately, due to a lack of sufficient individual effect sizes the potential relationship with violent recidivism could not be further explored. Fifth, only two studies reported effects for BPD, making it impossible to examine the risk-relevance of this condition in meta-analyses, despite frequent claims in the literature of its clear link with antisocial, interpersonally aggressive, and criminal behaviour (e.g., persons suffering BPD have been found more likely to commit serious violent offenses including acts of intimate partner violence [IPV]; Lawson et al., 2010; Newhill et al., 2009; Ross & Babcock, 2009). Personality disorder was examined as a unitary construct (i.e., BPD symptomology and personality disorders not otherwise specified [NOS]) but did not emerge as a significant predictor of recidivism. Moving on to non-diagnostic indicators, contrary to the hypothesis, existing theoretical accounts, and previous empirical findings, self-harm and suicidality did not significantly predict recidivism and the effect size was so small it is unlikely to have any clinical relevance. Lastly, psychiatric history was the strongest predictor of recidivism obtained in this

study and was in the range of the average effect reported in the field of psychology as a whole ( $d = 0.40$ ; Cumming & Calin-Jageman, 2017).

Taken together, findings suggest some aspects of women's mental health may be relevant to the prediction of recidivism. However, a crucial point to consider before delving into potential explanations for these findings is that the validity and reliability of a meta-analysis depends on the amount of data available and the quality of the included studies in terms of overall scientific and methodological rigor (Helmus & Babchishin, 2013). Thus, the current findings are best considered tentative and exploratory in nature; a necessary first step towards the construction of a solid scientific knowledge base. Identifying the processes that may underlie the relationship between mental disorder and the maintenance of criminal behaviour is challenging for a number of compelling reasons. The lack of an adequate understanding of the observed symptom clusters' underlying etiology (Ghaemi, 2018), as well as how internalizing symptoms and externalizing symptoms other than those considered antisocial may affect the initiation and maintenance of criminal behaviour makes causal inferences particularly difficult to make. Nevertheless, as mentioned in the Introduction to the current study, it is possible to establish some informed theoretical hypotheses, utilizing what is known about the potential effects of individual symptoms, particularly, those included as diagnostic criteria for multiple distinct disorders, on measurable behavioural outcomes. Thus, the following section will briefly address how some of these diagnostic criteria may relate to negative behavioural outcomes such as recidivism among adult women. I will address the two main conditions, depression and PTSD, any mental disorder as a singular construct, as well as the one general non-diagnostic indicator of mental health – psychiatric history – found to be significant predictors of recidivism in the sections below.

Depression emerged as a significant but modest predictor of recidivism. We know that symptoms of depressive disorders are associated with severe negative outcomes for women with these diagnoses, both in terms of subjective well-being and dysfunction, including behavioural dysfunction, in multiple areas of life. An understanding of how depression may negatively affect a person's ability to abstain from criminal behaviours can potentially be facilitated by an understanding of the determinants of personal change. Briefly, empirical evidence suggests that in order to change one's established behavioural patterns towards improved well-being, overall functioning, and better life outcomes, adequate motivation, energy, and an internal locus of control (i.e., a belief that change in one's external reality depends on one's own actions) are thought necessary. Consider a woman whose motivation is low, feelings of hopelessness intense, whose future appears beyond her own control and rather determined by uncontrollable, external factors alone, who is passified by apathy and plagued by suicidal ideation; is this woman likely to be immediately capable of successfully abandoning previous behaviours to which she has become accustomed, making serious, long-lasting changes in her life? Is any person- man or woman- able to relinquish dysfunctional or violent romantic connections and other relationships with antisocial others-who in many cases may be the only companions he or she feels share her reality and are close to him or her under such conditions? Is the abandonment of substance abuse a reasonable expectation of a woman whose baseline state is one of suffering? Will she seek and maintain the discipline necessary to successfully complete vocational training or education, abandoning antisocial systems of belief or world views in the face of the less than inspiring environment of the punitive correctional institution? Will she naturally experience an internal locus of control – against the backdrop of a lifetime of experiences, including traumatic events- which at that time were beyond her willful control- and believe in her ability to make the

changes necessary and further, that these changes will lead to be better life? When suicidal ideation takes hold, will she be able to see the purpose in these attempts at change? When one employs empathy to the lived experiences and realities of women- let alone women incarcerated for more than two years- it seems altogether less surprising that recidivism rates are slow to change.

Recall that PTSD was significantly associated with recidivism but only modestly so. Individuals suffering from PTSD experience stress-induced negative affect, including clinically significant anger and hostility, chronic hypervigilance, paranoid ideation, acute fear, flashbacks to traumatic events that in many cases involved acts of extreme violence and other personal attacks such as sexual assault (Donley et al., 2012; Jäggi et al., 2016). Not infrequently do these intense emotional experiences result in interpersonally violent responses to the perceived environmental threats. Systems of meaning and world views altered by the traumatic event(s) further entrench beliefs about the world as unjust, dangerous, and acutely threatening, necessitating a proactive, self-protective, and sometimes aggressive approach (Harris, 1993; Jäggi et al., 2016; Jernigan, 2000; Kraska & Kappeler, 1997; Moore & Elkavich, 2008; Parker et al., 2010). Moreover, it has been found that a proportion of trauma-victims and individuals diagnosed with PTSD develop an addiction-like response to trauma causing them to become involved in reenactments of the traumatic experience and risky, antisocial behaviours in general (see Levy, 1998 for further elaborations of several established theoretical accounts that seek to explain this effect). Dissociation is not uncommon in PTSD and other trauma-related mental health conditions and coupled with unpredictable attacks of fear and anger, aggressive and violent behaviours can occur. This psychiatric presentation is further combined with a chronic sense of a foreshortened future, meaning that the client perceives her life expectancy to be short,

which naturally encourages and facilitates impulsive behaviour and short-term life strategies, including criminal behaviours (Herman, 1992; van der Kolk, 2001). Lastly, empirical research findings to date suggest that the experience of certain PTSD symptoms such as nightmares and flashbacks is associated with the use of illicit substances (Chilcoat & Breslau, 1998; Cornelius et al., 2010; Duncan, 1974; Jäggi et al., 2016; Rich & Grey, 2005; Slade et al., 2008). The combination of these multiple, inter-related problems resulting from traumatic experiences and evidenced in PTSD and other trauma-related conditions are theoretically linked to engagement in impulsive, risky, antisocial, self-destructive, and both non-violent (e.g., drug offenses; Jäggi et al., 2016) as well as violent crime (e.g., Komaroykaya, 2009; Beckham et al., 2000).

Next, the unexpected finding that any mental disorder predicted recidivism, producing the largest significant effect size obtained in the study, deserves mention. There are multiple potential explanations for this finding. First, it may simply reflect that a disproportionate number of the women included in these samples were diagnosed with the most risk-relevant disorders included in each study's definition of 'any mental disorder'. Only a minority of studies reported the proportions of their sample meeting diagnostic criteria for each disorder, if indeed the disorders included in the definition of any mental disorder were reported at all. Conversely, it could mean that many women experienced the symptom clusters common to many diagnoses typically included in these studies; 1) lack of motivation, hopelessness and suicidality, an external locus of control, and low self-efficacy (e.g., MDD, PTSD), 2) emotional instability, particularly dysregulated anger (MDD, BPD, Bipolar disorder), fear (PTSD), hostility and feelings of dissociation and social alienation (MDD; PTSD), lack of trust in themselves and others, a foreshortened future, devaluation of one's own life, and dangerous world beliefs that together support impulsive, short-term life strategies that include substance abuse and risk-taking

(PTSD, BPD), and 3) impaired empathy (ADHD; BPD; Bipolar disorders), impulsiveness (ADHD; BPD, PTSD, Bipolar disorders), and a sense of entitlement (Bipolar disorders [‘mania’], BPD). Finally, an alternative interpretation is that simply being afflicted with any mental disorder increases a woman’s risk of recidivism through some mechanism that is not yet understood.

Lastly, psychiatric history predicted general recidivism; analyses resulted in an effect size in the modest to moderate range. Having a psychiatric history was associated with increased recidivism. Individuals with more severe psychiatric disorders, particularly those that involve externalizing symptoms and other behavioural indicators easily identified by laypersons in the community, are more likely to make contact with psychiatric services both in the community and while incarcerated. For example, a person suffering acute psychosis or a manic episode who is behaving irrationally, disturbingly (e.g., urinating in the garbage can inside the local gas station in an unprecedented fashion while loudly responding to [likely] auditory hallucinations), or even threateningly in public is more likely to be identified as having a mental health problem and thus to be addressed psychiatrically in some fashion relative to an individual with mild generalized anxiety without overt, externalizing symptomology). Thus, this finding appears to support the contention that severe mental health challenges, perhaps particularly those with externalizing features, may be at increased risk of recidivating.

The finding that race moderated the relationships between disorders and recidivism (i.e., depression and anxiety), such that significant positive associations appeared limited to samples consisting primarily of White women is difficult to account for using existing theory or empirical findings reported in psychological and correctional research. It seems plausible that the potential causes underlying these effects may be found in the systematic racial bias that has been

consistently found in assessment and diagnoses of psychiatric disorders<sup>17</sup>, including in correctional populations (e.g., Baglivio et al., 2016), sentencing disparities along racial lines (Skeem & Lowenkamp, 2016), or the intersection between the two. However, a sound empirically supported theoretical account of the exact factors, underlying mechanisms, and processes that may interact to create this relationship is not yet possible to establish. This is difficult in part because the degree to which observed differences in diagnostic rates relate to the *accuracy* of the psychiatric assessments conducted in forensic contexts or the *validity* of sentencing decisions in terms of adherence to current legislation and legal precedence based on case law, remain unknown. Practically, whether the diagnostic rates among Whites more accurately represent actual prevalence rates of disorders or whether the rates found among, for example Blacks, should be considered more accurate, has thus far not been possible to establish. Such evaluations would be tautological in that they would necessarily depend largely on the assessments themselves, which may well be the original source of the systematic disparities in the first place. Further, we know that psychiatric diagnoses are simple descriptive terms and not scientifically derived constructs; the etiology of most mental conditions remain largely unknown as do optimal treatment approaches. Thus, attempts to determine what constitutes accurate diagnosing of the underlying latent dysfunction are further complicated. In summary, in order to address questions regarding explanatory factors involved in racial differences of the relevance of various mental disorders to the prediction of recidivism depends on a long-term program of well-designed research studies by researchers dedicated to maximizing adherence to scientific principles and methodological and reporting standards. First, diagnostic criteria better informed

---

<sup>17</sup> Note that such biases have also been reported to occur in correctional risk/needs assessments (Skeem & Lowenkamp, 2016).

by etiological factors that are likely to uncover the true prevalence rates in the population must be established. Second, the overall accuracy of diagnostic systems must be examined against the backdrop of these prevalence rates. Of course, achieving the prevalence rates without relying on inaccurate diagnoses is a difficult problem to solve in and of itself. Third, the reasons for sentencing disparities across race must be understood, and lastly, the interaction of all of these elements must be thoroughly examined. Together these factors would finally be used to understand this complex issue.

Finally, a single null finding is also worth discussing briefly among the major findings. An a-priori hypothesis was made stating that anxiety was expected to emerge as a protective factor against recidivism. The theoretical grounds of this hypothesis was simply the idea that engaging in certain crimes might be expected to require a certain degree of risk and distress tolerance, as well as general calm. It was thus theorized that given that individuals suffering non-negligible anxiety are more likely to avoid potentially anxiety-inducing activities, an individual is suffering clinically significant anxiety may be less likely to engage in criminal behaviour. Despite its apparent face validity, this hypothesis was not supported in the current meta-analysis. Instead, null findings were obtained; neither a positive nor a negative relationship with recidivism was supported. Thus, consistent with results of some previous empirical research (e.g., Hubbard and Pratt's (2002) meta-analysis looking at the effect of anxiety on recidivism among delinquent girls), the current meta-analysis suggests that anxiety may be among the less risk-relevant mental health conditions.

### **Limitations**

Although the current meta-analysis is not beyond criticism in terms of its internal methodology and analyses, every effort was made to maximize the information gleaned from the

literature and employ the techniques and standards most likely to produce valid and reliable results. However, the studies included in the meta-analysis, and thus by extension, the synthesis itself, were subject to several non-negligible limitations, each one of which will be addressed in the sections below.

First, the number of included studies was small, making conclusions tentative and findings in need of replication in future research. Second, and most striking among the limitations uncovered, was the fact that no studies accounted for diagnostic comorbidity. Thus, in any given study, the inclusion of a woman in one disorder group (e.g., general anxiety) did not preclude her from a) inclusion in another disorder group (e.g., depression), and b) in the overall comparison group consisting of all the women who did not have the disorder of interest in the given individual analysis. Thus, resulting comparison groups likely consisted of a large proportion of women with another disorder other than the one predicting recidivism in any given analysis. Depending on the nature of the predictor diagnosis as well as the nature of the diagnoses of the women in the comparison group, this could cause attenuated or artificially inflated effect sizes. In short, no studies ensured that the comparison group consisted of women without mental disorder. The fact that most studies only assessed a small proportion of the most common disorders further exacerbated the potential inferential problem resulting from this limitation.

As mentioned in the introduction to the literature above, this was a main limitation of another major meta-analysis conducted on primarily mentally disordered justice-involved men (Bonta et al., 2014). Considering that recent estimations of the proportion of federally incarcerated women in Canada diagnosed with at least one mental disorder, which was approximately 80% of the total population in 2018 (OCI, 2019), adequately addressing this

limitation would be extremely challenging at the recruitment level. Nevertheless, a minimum potentially achievable standard would be to attempt to exclude women who have disorders strongly suspected of increasing the risk of recidivism, such as BPD, from the non-disordered comparison group.

Another considerable limitation was the fact that none of the included studies ensured group equivalency by using techniques such as matching procedures. Even more surprising, no studies conducted post-hoc analyses examining differences among disordered groups in terms of extraneous risk-relevant constructs. Thus, the disordered and non-disordered groups could have differed significantly on important factors including the main gender-neutral risk factors (i.e., the central eight, including factors such as antisocial history, antisocial associates, and substance abuse) and overall estimated risk of recidivism. Thus, the possibility remains that any between-group differences found could have been accounted for by relevant factors other than the mental health predictors under investigation. On a related note, only a single study reported average or majority risk classifications for the overall sample, making investigations into the risk-relevance of mental health status by risk level impossible.

Next, a number of related limitations to sampling and reporting standards were uncovered. First, the proportion of the women who had recently received mental health treatment specifically suited to their disorder either prior to sentencing, while serving the sentence in custody or in the community, or sometime during the follow-up period at risk in the community, was not reported in a single study. Second, the proportion of women prescribed psychopharmaceuticals for their mental health condition and the associated medication compliance rates were never reported. Third, the proportion of the overall sample who met diagnostic criteria for the disorder under investigation in recidivism analyses was often not

reported. For example, for self-harm and suicidality, only one out of the four included studies reported the proportion of the sample that had attempted suicide in the past, the obtained correlation coefficients from the remaining studies were reported without prevalence rates (e.g., Olson, 2016). If the prevalence of the indicator was low in an already limited sample, particularly if the recidivism base rate was also low, the implications for statistical power to detect significant effects would be considerable.

Shifting focus to methodology and measurement limitations, it is worth noting that definitions of both predictors and outcome variables were wholly inconsistent across studies. For example, for the any mental disorder predictor, three studies failed to report the diagnoses included in their ‘catch all’ variable, and two studies reported disparaging and narrow definitions. Similarly, length of follow-up and most importantly, base rates, varied immensely. Recidivism rates ranged from 7% to 59%. Minimum recommended base rates for recidivism is 10%, and women-centered researchers suggest it should be higher when the population under study is justice-involved adult women (Tabachnick & Fidell, 2019). In summary, the variability in all aspects of methodology and measurement made it difficult to determine whether the sampling reflect the same population of effect sizes or even the effect sizes for the same population of women.

### **Future Research**

First and foremost future research should be conducted in order to accumulate a larger number of viable effect sizes for both the predictors included in the current study as well as predictors for which virtually no data yet exist and that were excluded from the current syntheses. In particular, researchers should further investigate the risk-relevance of PTSD, especially as it pertains to violent recidivism, ADHD, and bipolar disorders. Perhaps most

relevant is BPD, which has substantial theoretical and empirical evidence supporting its potential relevance to recidivism risk. Further, despite the fact that no meaningful relationship between self-injurious behaviour and suicide attempts and recidivism was uncovered in the current study, future investigations should continue to investigate the correlates and distal outcomes related to these externalizing behaviours.

Researchers attempting to study the relationship between mental health and correctional outcomes among adult women in future studies should make every effort to address the limitations of existing research and adopt higher standards in terms of sampling procedures, overall methodology, analytical approaches, and reporting practices. Reporting practices left much to be desired; researchers must make efforts to report crucial contextual information that could reasonably be expected to have a major impact on resulting effect sizes. For example, the proportion of the sample with the mental disorder or indicator of interest as well as the proportions of women with other potentially relevant disorders as well as the disordered subsample's mental health treatment status and medication compliance rates need to be reported. The two latter variables are among the most crucial in this research context as both could be expected to partly or completely extinguish any effect associated with the disorder of interest, if found effective. Lastly, a major requirement of future research would be to assess the estimated risk of the sample and the potential presence and effect of traditional gender-neutral factors among women in their respective samples. Without addressing established risk-relevant constructs alongside mental health, the utility of assessing mental disorder and other indicators of mental health status, and using these to predict recidivism, will remain somewhat unclear.

In future studies, researchers should aim to examine a wider range of correctional outcomes, including violent recidivism specifically and technical violations of the terms of

conditional release. Further, empirical research to date suggests that women with severe mental health challenges are disproportionately involved in serious institutional infractions, including violence, and represent a continuous challenge to institutional management and control, as well as to both staff and clients' safety and well-being (Lord, 2008). These women are typically met with punitive measure that further exacerbate these aspects of prison life and culture (Haney, 2001; Lord, 2008) and that may undermine therapeutic effects of correctional programming. Thus, further empirical investigations are also required to examine institutional outcomes for mentally ill women, in terms of prevalence rates, mental health needs profiles, and the viability and efficacy of developing alternative management strategies that minimize the use of punitive measures that exacerbate existing mental health conditions while maintaining institutional order.

Next, the current finding that the predictive utility of psychiatric diagnoses depended on the race of the women assessed is likely to require long-term, scientifically rigorous, interdisciplinary research across psychiatry, correctional psychology, and law, in order to begin to assess the potential causes and correlates of racial bias consistently observed in all three areas. The success of such efforts, in turn, will depend in part on the establishment of a scientific basis for the diagnoses of various common mental disorders, which remains a fundamental problem within the field of psychiatry.

The current meta-analysis focused on examining potentially direct associations between mental health status and recidivism among adult women. Nevertheless, one final discovery made as a result of the extremely extensive literature review conducted is the virtual absence of any empirical research examining mental health as a potential responsivity issue. In the process involving the identification over 9, 000 records resulting from the search of a large number of academic and alternative sources in multiple languages using an extremely large number of

search terms, virtually no studies focusing on this issue were detected. Specifically, not a single study discovered had attempted to compare treatment change and post-treatment recidivism rates among disordered and non-disordered women. This finding was extremely surprising, given that both gender-neutral and gender-responsive scholars agree that regardless of any direct risk-relevance of mental health status among women, most current mental disorders would be expected to have a measureable impact on women's ability to fully participate in and benefit from, existing correctional treatment interventions; intuitively, it would seem that struggling with severe mental health concerns could have the potential to seriously reduce any positive effect of treatments on women's ability to succeed in the community in the long term.

### **A Note on Practice and Policy**

The current meta-analysis, despite its wide scope and utilization of methods and analyses most likely to result in reliable and valid results, should be interpreted a first, exploratory step in a longer-term program to research dedicated to the mental health of justice-involved adult women. Thus, making definite practice and policy recommendations based on this study alone would be inappropriate. Nevertheless, it is worth mentioning that correctional agencies routinely conduct their own treatment outcome and efficacy studies, and given that the current findings suggest, across several analyses and across moderators, that some mental health conditions are related to recidivism, these agencies and their clients would likely benefit from methodologically rigorous treatment change and its relationship to correctional outcomes for both disordered and non-disordered women separately. This would constitute a starting point for the development of either pre-treatment mental health interventions for those so acutely affected that their participation in ongoing programming is unlikely to illicit any change (i.e., specific responsivity), and/or develop additional treatment modules that specifically target symptom

clusters linked directly with increased recivism. In addition, correctional authorities are mandated to provide safe and humane living conditions for all incarcerated clients, and this includes women with mental illness. Further, mental illness affects day-to-day institutional operations negatively. It would therefore be highly recommended that such agencies make targeted efforts to gain a better overview and understanding of the needs of their various populations – most especially adult women. Further elaborations on practice and policy recommendations follow both in the Discussion section for Chapter 3, which details a latent class analysis (LCA) of the mental health profiles of a large sample of incarcerated adult women, and in Chapter 4, which constitutes a summary of findings, limitations, and recommendations based on the current program of study as a whole (i.e., Study 1- the meta-analysis- and Study 2 – the latent class analysis).

### **Summary**

The current study constitutes the first quantitative synthesis of existing empirical findings regarding the relevance of mental health to justice-involved adult women's correctional outcomes. The findings provide crucial support for the construction of a scientifically sound, evidence-based understanding of the needs of this understudied population, provide rich data that can help guide and improve the quality of future studies, and puts mental health on the map as a relevant factor to consider in the management and treatment of adult women in conflict with the law. The results speak to the gender-neutral vs. gender-responsive debate in so far as it supports the contention that aspects of mental health are relevant factors to consider in women's corrections. More research is needed however, to facilitate an understanding of the exact nature of this relationship and how it may best be addressed. Despite the need for replication, as well as further research with a broader scope in terms of the mental health indicators and outcome types

examined, the study findings lend further support for gender-responsive scholars' and other experts in the area of women's corrections existing recommendations; future correctional practice should be guided by gender-responsive principles involving holistic, mental health- and trauma-informed approaches to the management and treatment of justice-involved adult women.

### Chapter 3

#### **Study 2: Mental Health Profiles of Justice-involved Adult Women: A LCA of the SPIn-W**

This chapter presents Study 2 of the two-study program of research, a latent class analyses (LCA) focused on the mental health profiles of a large sample of incarcerated adult women in Maine. The study examined whether and how women can be grouped into distinct classes based on their respective mental health challenges or absence thereof. The purpose of this endeavor is to inform and to contribute to the empirical literature and our overall understanding of justice-involved women's mental health status and how it may relate to outcomes such as recidivism.

The chapter begins with the presentation of the findings of previous studies; first, the contribution of early feminist scholars focused on person-centered, qualitative and quantitative analyses of the different pathways to crime for justice-involved women are described. Second, a presentation of LCA as an analytic strategy is discussed in order to facilitate a full understanding of the research findings resulting from previous studies using this technique; the findings of recent LCA studies in which the mental health profiles (i.e., classes) of adult women were the primary focus are then presented. Third, a summary focused on understanding the mutually informative relation between early pathways work and recent LCA findings, ending with tentative take-home messages is presented. Fourth, a brief summary of previous findings that speak to the potential role of childhood adversity and its contribution to mental health challenges in adulthood are discussed. Childhood adversity is worth examining briefly here, as it is an extraneous factor (i.e., an auxiliary variable) to be included in the mental-health focused LCA of the current study. Next, the current LCA is described and discussed in its entirety, concluding

with potential implications and recommendations for future research, correctional practice, and policy.

Gender-responsive research started in the late 80s and early 90s as a response to the existing traditional gender-neutral scholarship's exclusive focus on justice-involved men. This pioneering work was driven by a small number of feminist scholars who primarily utilized qualitative, person-centered analyses in order to understand women's pathways into criminal behaviour (Chesney-Lind, 1989). The reliance on qualitative and person-centered, rather than the quantitative, variable-centered approaches utilized in traditional, men-centered correctional research later came under scrutiny and vigorous criticism by advocates of the gender-neutral approach. Proponents of the gender-neutral perspective argued that the qualitative techniques and other methodologies such as case studies did not meet the standards of scientific rigor and could result in non-evidence based practice (Van Voorhis et al., 2010; Wright et al., 2007).

The gender-responsive research literature has grown exponentially since then, however. Today, women-centered research, extensions of the early feminist research are used to inform research using more quantitative methods of investigation than the early feminist literature did (e.g., Van Voorhis et al., 2010 vs. Chesney-Lind, 1989).

Early pathways models used women's own narratives to retrospectively identify multiple distinct pathways into crime (Jones et al., 2014), linking factors such as childhood adversity and victimization, mental health challenges such as depression, anxiety, and substance use to offending (Covington, 1998; Daly, 1992). These early models of factors and pathways into crime have been replicated and refined in more recent empirical research employing quantitative pathways analyses. For example, distinct pathways characterized by victimization, socially withdrawal, and depression (Covington, 1998; Daly 1992) overlap substantially with pathways

identified in recent empirical research (Brennan et al., 2012; Salisbury and Van Voorhis, 2009). More recently, gender responsive researchers have turned to person-centered, complex latent variable modeling techniques such as LCA or LPA in efforts to determine whether the factors that emerged in early qualitative and later quantitative pathways studies, as well as more contemporary research employing alternative quantitative modeling methods would replicate. Replication of previous findings would support the contention that there may be risk factors specific to or more salient among women than among men, supporting the gender-responsive perspective.

Latent class and latent profile analyses (LCA; LPA) as statistical modeling techniques are presented briefly below, to facilitate a better understanding of the findings of gender-responsive research studies employing these techniques. Next, the most relevant latent modeling research conducted in the past 15 years is presented.

### **Fundamentals of LCA and LPA: A brief Overview**

LCA/LPA have been utilized in contemporary research to examine typologies of justice-involved women. It is important to understand that LCA and LPA are based on the same fundamental principles of person-centered, latent modeling. Although there are differences in analyses at the micro-level (e.g., the specific type of algorithm employed to facilitate model estimation), the only relevant difference is the type of data used; binary variables are used in LCA while LPA allows for the inclusion of continuous variables.

The purpose of LCA and LPA is to classify individuals into the same or different groups (i.e., classes or profiles) depending on the degree of similarity between the individuals on a number of observed variables called indicators (Tein et al., 2013). In the case of LCA, individual women's binary scores on variables such as psychiatric diagnoses are entered as indicators into

the latent modeling environment. Next, the researcher specifies the class solution; the number of classes hypothesized to be required to account for all distinct, latent response patterns present in the sample. The resulting estimated model is subsequently evaluated in terms of how well the proposed classes account for the data according to a series of objective statistical, theoretical, and inferential criteria. Women whose response patterns are similar across indicators are considered part of the same latent class. The proportion of the sample that makes up each class, and the estimated probability any individual woman belonging to a class has of endorsing (or not endorsing) each individual indicator are produced.

There are a number of advantages to these techniques when applied to heterogeneous populations (Turner et al., 2008). First, they are person- rather than variable-centered, allowing for potentially holistic evaluations of the individuals themselves. Second, the models allow for a wide range of phenomena to be examined together. Third, LPA can accommodate variables with different measurement levels, all within the same analysis (e.g., both binary and continuous indicators). Third, LCA includes all participants' data, even if incomplete, without the missing data having undue effect on the accuracy of classifications or the viability of the model itself (Turner et al., 2008). Fourth, latent variables reduce measurement error relative to observed variables, again supporting precise estimates and statistical indices (Turner et al., 2008). Fifth, LCA allows for the construction of structural models that include contextual variables that help characterize the latent classes. For example, to improve model fit and classification accuracy- a covariate- an extraneous variable that is believed to have a causal inference on the covariation between individual indicators- can be included (Turner et al., 2008). Sixth, relative to standard cluster analysis, LCA also has the added benefits of being data-driven, relying less on the researcher's personal judgment (Campbell et al., 2019). Lastly, analyses produce a probability

statistic of a person belonging to a given class as well as probability estimates of any individual woman grouped into a given class endorsing each indicator (Herbert et al., 2007).

### **Previous LCA and LPA: Mental Health Profiles of Justice-involved Women**

Below I explore individual study findings of previously conducted LCA and LPA (see Table 21 for a summary of results of these studies). This section concludes with an overall summary of findings and identifies knowledge gaps in the existing latent modeling literature pertaining to the mental health profiles of justice-involved women.

Herbert et al. (2007) examined latent classes in a sample of 149 incarcerated women in the U.S. Indicators in the models were trauma-related (i.e., crime, disaster, sexual, and physical trauma). Psychological functioning was defined as the presence or absence of depression, PTSD, and generalized anxiety disorder (GAD), and was included in the study as a validation measure. A two-class solution best fit the data. While Class 1 did not evidence elevated scores on the depression, PTSD, or GAD, Class 2 ( $n = 72$ ) showed significant elevations on all three; depression ( $p = .02$ ), PTSD ( $p = .001$ ), and GAD ( $p < .001$ ). These women were also more likely to have experienced all types of trauma relative to women in Class 1.

Welch (2007) examined latent profiles of adjudicated girls ( $N = 203$ ) in North Carolina according to a number of social and psychological variables. Psychological variables included depression, aggression, self-esteem, personal stress, guilt, suicidal thoughts, confused thinking, disturbing thoughts, memory loss, and alcohol and drug use. In addition, the author examined problems with parents, friends, and school, and educational status. LPA revealed that a four-class solution best fit the data. Class 1 “Aggression only” (51%) was characterized by elevated scores on aggression and subclinical scores on depression, problems with family, peers, and school, and drug and alcohol abuse.

**Table 21**

*Latent Class Solutions (Number of Classes) in Previous LCA and LPA*

Study	N	Sample	Classes/ Profiles	Mental health indicators
Herbert et al. (2007)	147	Women	C1 C2	Low needs <b>Elevated needs: Depression, PTSD, GAD</b>
Joosen et al. (2016)	397	Women	C1 C2 C3 C4	Low probability of prior treatment Low probability of prior treatment <b>High probability of prior treatment</b> Low probability of prior treatment
Perkins (2010)	765	Women	C1 C2 C3 C4	Limited/no psychiatric history, suicide Limited/no psychiatric history, suicide <i>Moderate use of psychiatric services, poor coping</i> <b>High needs: Suicide attempts (50%), increased use of psychiatric services</b>
Turner et al. (2008)	90	Women	P1 P2 P3	Low needs <i>Risk of anxiety, depression, paranoia, schizophrenia, BPD</i> <b>Clinically significant somatization, anxiety, depression, paranoia, schizophrenia, BPD</b>
Wanamaker (2020)	1,684	Women	P1 P2 P3	Low needs: few adverse childhood experiences <i>Moderate needs: more adverse childhood experiences, low aggression</i> <b>High needs: more adverse childhood experiences, elevated aggression</b>
Brown et al. (2021)	100	Girls	P1 P2	<b>Complex trauma</b> Low overall needs
Guthrie et al. (2012)	153	Girls	P1 P2 P3	Low depression, anxiety <b>High depression, anxiety</b> <i>Average depression, anxiety</i>
Wagstaff (2020)	354	Girls	P1 P2 P3	Low needs <i>Moderate needs</i> <b>High needs</b>
Welch (2007)	203	Girls	P1 P2 P3 P4	Low depression <b>High depression</b> <b>High depression</b> <b>High Depression</b>

*Note.* Results of studies utilizing samples of delinquent girls are highlighted grey and presented in the bottom table rows. Classes/Profiles with moderate or average mental health challenges are italicized; those demonstrating the most severe/greatest mental health problems are **bolded**.

Class 2 “Aggression and drug use” (18%) had elevated scores on aggression, depression, problems with school, and alcohol and drug abuse. Class 3 “Severe alcohol and drug use” (5%) had clinically elevated scores on depression, family problems, and problems with school, as well as extremely elevated scores on alcohol and drug use. Finally, Class 4 “Family conflict” (26%) had clinically elevated scores on depression, family problems, problems with school, and aggression, as well as subclinical scores on alcohol and drug abuse. In summary, three classes emerged as having elevated scores on one psychological variable, namely depression.

In a sample of women sexual offenders ( $N = 79$ ), Turner and colleagues (2008) examined latent profiles informed by criminal histories, personality indicators, and mental health variables. The Personality Assessment Inventory (PAI; Morey 1991) was used to assess mental health. The authors selected these variables in an attempt to fill a gap in the LCA literature for women sex offenders. The criminal histories were included in LCA in order to remain consistent with previous LCA of women sex offenders. The personality and mental health variables were included in order to address what the authors perceived as an important gap in the existing literature. The criminal history variables and the personality and mental health variables were addressed in a separate LPA. A three-class solution best fit the data. Class 1, called the “elevated drug and alcohol use” group, yielded scores that fell below the cutoff points to meet diagnostic criteria in most areas. However, women in this class had scores on alcohol and drug abuse that approached clinical significance. Class 2 ( $n = 39$ ), the “moderate psychopathology” group, had scores on the anxiety and anxiety-related disorders, depression, paranoia, schizophrenia, and borderline personality disorder scales that fell in the ‘at-risk’ range. Cutoff points for other areas fell below the threshold for clinical significance. Class 3, a very small class ( $n = 10$ ), yielded clinically significant scores in the areas of somatization, anxiety, depression, paranoia,

schizophrenia, and BPD. Women in Class 3 were also characterized by lower levels of alcohol abuse than Classes 1 and 2, and had similar drug abuse scores to Class 2.

Joosen and colleagues (2016) conducted a LCA for women ( $n = 397$ ) and men ( $n = 1,904$ ) incarcerated in the Netherlands. Analyses were conducted separately by gender. The following indicators were used to classify the women into groups: age of onset of offending, offense type, family situation in childhood, substance use, homelessness, economic marginalization, romantic relationships, parenthood, and mental health. Although a precise justification for the selection of these specific variables was not given in the study, two broad reasons were given for the selection: (a) findings of studies of women offenders suggest that these offenders have experienced a number of negative life circumstances, including unstable housing, violent relationships, financial difficulties, substance abuse, and mental health problems, making these variables particularly relevant to LCA and (b) the nature of contemporary research practices; Dutch researchers are investigating the life histories, criminal trajectories, and carceral experiences of women offenders. Mental health was broadly defined using a dichotomous variable indicating whether the women had received any mental health treatment for depression, anxiety disorder or phobia, alcohol or drug addiction, ADHD, psychotic disorder, personality disorder, or any other emotional or psychological problems in the year prior to their arrest. A four-class model best fit the data. While Classes 1, 2, and 4 were not characterized by women with mental health problems, Class 3, known as the “multi-problem property offenders” ( $n = 65$ ; 16.4%), was characterized by women who had a multitude of problems in several domains. These women were economically marginalized, and reported parental deviance, early first drug use, running away in childhood, soft and hard drug use in the year prior to the arrest for their index offense (i.e., the offense for which they were incarcerated

at the time of the study), and had also been treated for mental health problems. Moreover, these women were likely to be in a dysfunctional relationship, to be unemployed parents, and to have significant debts. Interestingly, among men offenders, a class that was very similar to Class 3 among women offenders emerged.

Perkins (2010) examined latent classes of women offenders ( $n = 765$ ) according to gender-neutral dynamic risk factors (employment, marital/family history and relations, associates/social interaction, substance abuse, community functioning, personal/emotional [i.e., mental health], and attitudes) and an indicator of suicide risk. A four-class solution was determined to be the best fit for the data. Women offenders in Classes 1 and 2 had low probabilities of a mental health history and previous suicide attempts. However, Class 3, called the “poor mental health and coping” class (10%), was characterized by high probabilities of reporting the use of psychiatric in/outpatient services (.72) and poor coping with stress (.76) as well as low probabilities of associating with substance abusers (.01) and unstable accommodations (.27). Class 4 (27%), the “overall high need” group, showed a high probability of poor coping with stress, and probabilities above .50 for suicide attempts (.51) and receiving in/outpatient services in the past (.66). High probabilities were also observed for having negative relations with parents in childhood, abusing alcohol or drugs, and associating with substance abusers. Probabilities above .50 were present in Class 4 for unstable accommodations, lower level of education, parenting responsibilities, and easily influenced by others.

Guthrie et al. (2012) examined differences in female juvenile offenders ( $n = 153$ ) using psychological functioning, risk behaviour engagement, and health status. LPA was used. The selected variables were chosen in order to address the lack of LCA and LPA research on comprehensive health among adjudicated girls. Depression and anxiety symptoms were assessed.

A three-profile solution was determined to be the best fitting model. Profiles 1 and 3 did not have elevated scores on depression or anxiety. Profile 1, identified as “low to moderate health risk” ( $n = 35$ , 22.9%), reported lower mental health problems, while Profile 3, the “compound health risk” profile ( $n = 50$ , 32.7%), reported scores close to the mean for the sample. Profile 2 ( $n = 68$ , 44.4%) showed both elevated depression and elevated anxiety scores, and was named the “high mental health risk” profile. This profile was associated with moderate health risk behaviours, lower subjective health ratings, and fewer daily hassles. Girls in Profile 2 reported committing fewer crimes than girls in Profiles 1 and 3 and more maternal support and warmth, as well as more neighborhood support.

Brown et al. (2021) examined the criminogenic needs and aspects of mental health associated with complex PTSD in a sample of 100 justice-involved girls between the ages of 12 and 21, the majority of whom had committed serious offenses. Some girls were in custody while others were serving community sentences. Complex PTSD is an emerging psychiatric diagnosis that includes broadly assessed mental health and psychosocial symptoms that are believed to result from chronic, varied, and severe trauma (van der Kolk, 2001). The ICD-11 (WHO, 2018) was used to measure complex PTSD. The most relevant items on the complex PTSD measure included PTSD scores, ACES scores, and scores on emotional dysregulation, attention dysregulation, substance abuse, and psychopathy. Two profiles emerged. Girls in Profile 1 were found to have complex needs; the average PTSD was more than double that obtained for Profile 2 – the low overall needs group. Further, girls within Profile 1 had over six adverse childhood experiences on average, relative to four for the low needs Profile. Scores on emotional dysregulation, substance misuse, and psychopathy were also higher. Lastly, this group had slightly higher scores on criminogenic needs (e.g., criminal associates), relative to Profile 2 girls.

Wagstaff (2020) conducted an LPA study with a sample of 354 girls between the ages of seven and 23 receiving correctional services in the community in Milwaukee County (U.S.). The latent profiles were based on the indicators derived from risk/need assessment designed to estimate the risk of recidivism of youth in conflict with the law; the Youth Assessment Screening Instrument (YASI; Orbis Partners, 2000). Mental health indicators included a mental health flag variable. The mental health flag indicated whether and to what extent mental health concerns were noted in the individual's full risk assessment on a 3-point Likert-type scale. Higher scores indicated more serious mental health concerns. Further, a measure of potentially traumatic experiences in childhood—a proxy measure of the original Adverse Childhood Experiences Scale constructed from individual items across multiple domains of the YASI (ACEs) was also included. ACEs consists of 10 binary questions related to a variety of adverse (i.e., potentially traumatic) experiences in childhood, with scores ranging from zero to ten. Three profiles emerged; Profile 1 “low needs” was characterized by low mental health needs (Mental health flag;  $M = 0.30$ ,  $SD = 0.68$ ) and few adverse childhood experiences ( $M = 0.63$ ,  $SD = 0.94$ ). Profile 2 “moderate needs” had moderate mental health concerns as indicated by the mental health flag ( $M = 0.78$ ,  $SD = 0.85$ ) and moderate scores on ACEs ( $M = 1.63$ ,  $SD = 1.59$ ). Profile 3 “high needs” had high mental health needs ( $M = 1.17$ ,  $SD = 0.89$ ) and the highest scores on ACEs. It is worth mentioning that despite having the highest ACEs score, the mean score on this measure modest ( $M = 2.38$ ,  $SD = 1.83$ ).

Wanamaker (2020) conducted a latent profile analysis (LPA) study with a sample of 1,684 women on community supervision in Canada. The Service Planning Instrument (SPIn; Orbis Partners, 2003) was used to derive indicators. The majority of women were low risk.

The mental health flag derived from the SPIn is identical to that used in Wagstaff's (2020) study above. ACEs was also used to measure of adversity in childhood. A three profile solution best fit the data. Women fitting Profile 1 had low mental health needs and few ACEs. The women in this class had low risk on the aggression domain of the SPIn. Women fitting Profile 2 had moderate mental health needs and ACEs scores and were at low risk for aggressive behaviour. Women in Profile 3 had the highest average ACEs and mental health flag scores, scored higher on the aggression risk domain, and had the most extensive criminal histories.

Taken together, the LCA and LPA analyses above typically resulted in one or more low mental health needs and one high mental health needs class. Which group represented the largest proportion of women or girls varied between studies; some found that the severe mental health needs group only made up a small portion of the sample, while others found they made up the majority (Brown et al., 2021). Three studies also uncovered a moderate mental health needs group (Perkins, 2010; Turner et al., 2008; Wanamaker, 2020). No studies have examined latent (i.e., unobserved) classes of women offenders according only to mental health factors or attempted to assess mental health in a comprehensive manner; most studies include a small number of general mental health indicators. A minority have included adversity or trauma in childhood in their models. Typically, groups identified have not been further examined in terms of the relationship between class or profile membership and estimated risk of recidivism, institutional outcomes, or general or violent recidivism.

The following section will present a brief overview of trauma and adversity as it pertains to the mental health of justice-involved women before the current study is described in more detail.

### **Trauma and Adversity in Childhood among Justice-Involved Adult Women**

The vast majority of justice-involved women have been exposed to trauma both in childhood and in adulthood. Among women in federal custody in Canada, roughly 33% suffer PTSD (OCI, 2019) and in empirical research studies of justice-involved women in the United States rates as high as 50% have been found (Dehart et al., 2013). In Messina and Grella's (2006) study of 491 justice-involved adult women, 84.3% were found to have experienced at least one traumatic event and 21.2% had experienced five or more traumatic events in their lifetimes. In terms of events that involve abuse or neglect, the most common traumatic experience was sexual abuse; 45.1% had experienced this form of abuse (Messina & Grella, 2006).

Trauma is linked with the development of mental disorder. Theory and empirical evidence overwhelmingly supports childhood trauma and adversity as precursors to the development of mental health disorders and overall poor mental and emotional functioning in adolescence and adulthood in non-forensic populations (Scheffler, et al., 2020) including among justice-involved women (Dehart et al., 2013; Ford et al., 2012; Messina & Grella, 2005). For example, Messina and Grella (2005) examined the relationship between childhood traumatic events and mental health problems requiring treatment in adulthood 500 women participating in a correctional treatment program in California. Interview data revealed that the impact of childhood traumatic events on health, including mental health, is strong. Specifically, childhood trauma was associated with a 40% increase in the odds of requiring mental health treatment in adulthood. Further, adverse childhood events are strongly and cumulatively associated with criminal behaviour and offense history. Among justice-involved women, it has been found that the greater the number of adverse events experienced in childhood, are associated with younger

age at first arrest, first lockup, and first drug use, as well as increased number of arrests and sentences to custody (Messina et al., 2007).

Given the powerful link between trauma and mental health challenges and between trauma and negative correctional outcomes, gender-responsive scholars consider trauma a gender-responsive risk factor. For example, risk/need assessments such as the Service Planning Instrument for Women – SPIn-W (Orbis, 2006) and the Youth Assessment Screening Instrument (YASI; Orbis Partners, 2000) both include all items from one of the most commonly used self-report measures of childhood adversity with both criminal justice and non-forensic populations—the Adverse Childhood Experiences Scale (ACEs). Researchers employ a variety of versions of this measure, including the original self-report scale, further refined versions, and equivalent proxy scales constructed from existing youth and adult gender-responsive risk/need assessments (Wanamaker, 2020).

In conclusion, childhood adversity as well as adult victimization are linked to poor mental health and correctional outcomes, and are thus important to consider in investigations of the mental health profiles of justice-involved women.

As discussed in detail above, the existing empirical literature suggests that there are multiple distinct pathways to criminal behaviour for justice-involved women, some of which involve severe psychiatric problems, substance abuse, childhood adversity, and other forms of trauma. Previous LCA and LPA studies have found that the population of justice-involved women do not fit into a single mental health or correctional risk/need profile; with most findings supporting the presence of one profile characterized by severe mental health challenges, extensive trauma histories, and elevated risk. However, to the best of the current author's knowledge, no prior LCA or LPA studies with adult justice-involved women have included more

than a small number of mental health disorders and other indicators of poor mental health. The mental health indicators in these models have typically not been the main focus and mental health has never been extensively or comprehensively assessed. In addition, previous studies have not examined the potential association between the mental health indicators, or the derived latent mental health profiles and recidivism outcomes. The current study aimed to address this gap in the literature by widening the scope of mental health challenges assessed and examining the resulting latent classes in terms of women's success upon release from custody.

### **The Current Study**

The current study utilized latent class analyses (LCA) to investigate the latent mental health classes (i.e., profiles) and their potential association with recidivism in a custody sample of adult women in the state of Maine. The overarching goal of this set of analyses was to uncover distinct forensic profiles of women differing in terms of mental health needs; internalizing disorders and externalizing problems, childhood adversity, demographics characteristics, offense histories, and general recidivism. The profiles provide valuable information about the prevalence of severe mental health problems in this population, the mental health needs associated with different subgroups latent in the population of justice-involved adult women, and the relevance of these classes to recidivism. Identifying these underlying classes could ultimately aid not only in developing a better understanding of the population of women in custody, but also in the identification of specific mental health needs, the development of more effective case management and treatment plans, modes of treatment delivery, and risk assessment.

The current series of LCA study focused on mental health and trauma, utilizing items from the Mental Health Domain of the SPIn-W Risk/Need/ Case Management Tool (Orbis Partners, 2006). These items were the main variables of interest in this study. In addition, scores

on the SPIn-W Pre-Screen risk measure were used as a covariate in LCA models to better inform classification and improve model fit. A proxy-measure of the ACEs as well as key demographic characteristics (i.e., age and race) were utilized as auxiliary variables to gain an improved understanding of the women in each class (see Appendix K). Lastly, recidivism at 24-months follow-up was included as a distal outcome.

In order to address a number of novel research questions about the nature of women's correctional mental health profiles, a combination of preliminary analyses focused on overall sample characteristics, mental disorder prevalence rates, and childhood adversity, latent class analyses identifying the latent class solution underlying the observed data, and exploratory post-hoc analyses was employed.

### **Research Questions**

1. What proportion of women in the current sample have a) at least one current mental disorder diagnosis, b) externalizing problems such as non-lethal self-harm and suicidal ideation and past suicide attempts, and c) severe, complex mental health needs?
2. What class solution will most accurately represent the data? Put differently, how many latent classes account for the observed mental health data in the current sample? Relatedly, what characterizes the overall mental health status of each derived class?
3. Will the incorporation of the Pre-Screen risk measure as a model covariate result in better overall model fit and improved classification accuracy?<sup>18</sup>
4. The number and nature of latent classes could not be determined a-priori. Thus, very firm hypotheses regarding potential differences on auxiliary variables could not be justified.

However, because at least one class was expected to have severe mental health needs

---

<sup>18</sup> Note that all items in the Pre-Screen Spin-W measure are presented in Appendix J.

relative to remaining groups, and childhood trauma exposure predicts adult mental health problems (Dierkhising et al., 2013), I cautiously anticipated that the ACEs total scores would differentiate between the low and high mental health needs classes. No predictions were made with regards to any additional classes, age or race.

5. Will rates of general recidivism at 24-, 38-, 48-, and 60-months follow-up differ significantly between classes?

### **Hypotheses**

1. Based on previous research on federally incarcerated adult women's correctional profiles, I hypothesized that between 70% and 80% of women would have at least one current diagnosis of mental disorder (Office of the Correctional Investigator [OCI], 2019). Further, I expected to find that between 10% and 25% would have problems with self-harm and suicidal ideation and past suicide attempts (Corabian et al., 2013). Further, I tentatively hypothesized that about 15% to 20% of women would evidence severe mental health needs (Blanchette, 1996).
2. Based on the LCA and LPA studies of similar samples of adult justice-involved women in custody in the United States and Canada the data would be best represented by two to four classes. One or more classes were expected to have relatively low mental health needs and at least one class was expected to demonstrate severe, complex mental health concerns.
3. The Pre-Screen risk score was believed to be a predictor of and to have an influence on the formation of classes. Thus, I hypothesized that including the covariate would improve overall model fit, classification accuracy, and class separation, clarifying the viability of the initially identified latent structure.
4. The number and nature of latent classes could not be determined a-priori. Thus, very firm hypotheses regarding potential differences on auxiliary variables could not be justified.

However, because at least one class was expected to have severe mental health needs relative to any remaining groups, and exposure to adversity (i.e., potential trauma) in childhood predicts adult mental health problems (Dierhising et al., 2013), I cautiously anticipated that the ACEs total scores would differentiate between the low and high mental health needs classes. No predictions were made with regards to any additional classes, age, race.

5. There is limited empirical support for a link between internalizing disorders such as anxiety and mood disorders and recidivism among justice-involved women. However, externalizing behaviours such as self-harm, homicidal ideation, and a small number of specific mental disorders including PTSD and BPD have been more consistently linked with rule-violations, aggression, and criminal behaviour, including recidivism (Benda, 2005; Dehart et al., 2014; Kulka, Schlenger, Fairbank, Hough, Jordan, Marmar, & Weiss, 1990; Jackson, Sippel, Mota, Whalen, & Schumaker, 2015; Jäggi, Mezuk, Watkins, & Jackson, 2016; King et al., 2018; Moore, Tull, & Gratz, 2017; Van Voorhis et al., 2010). Given that classes are derived empirically during the modeling process however, one could not ascertain that the majority of women in any one class would necessarily suffer these mental health related behavioral challenges. Thus, the only hypothesis made was that if a sufficiently populated class of women emerged with very high rates of these problems, their recidivism rates would be significantly elevated compared to (the) other class(es).

## **Method**

### **Participants**

The archival dataset utilized in the current study was provided by Orbis Partners, Inc, a private correctional firm located in Ottawa, Ontario, Canada. The dataset consisted of a sample of 920 adult women who were incarcerated in a Maine State prison between 2009 and 2019. All

920 women had been assessed at intake into the institution and thus had valid scores on the full Service and Planning Instrument for Women (SPIn-W) risk/need/case management assessment tool. Table 22 below provides sample characteristics. In terms of demographic characteristics, the average age was 34.52 ( $SD = 8.67$ ), with the youngest woman being 20 and the oldest being 70 years of age. Almost 90% of the women were White. Non-violent property offenses and drug offenses were the most common index offenses; only about 14% had committed violent offenses. Recidivism rates varied naturally depending on length of follow-up from 9.7% at 24-months follow-up and 17.5% at 60-months follow-up. It is important to understand the underlying reason for these unusually low base rates (e.g., recent cross-state estimates of average recidivism rates among incarcerated women suggest that approximately 30% are returned to custody over a three-year follow-up period; Deschenes et al., 2007). Recidivism was narrowly defined as any return to custody to a Maine state prison. Maine prison authorities did not have access to information regarding technical violations, new offenses, or subsequent returns to custody in other states or at the local level (e.g., incarceration in a local or remand jail). Thus, only women who returned to custody in a Maine state prison were considered recidivists.

Table 22 below provides a summary of basic sample characteristics for the initial sample of 920 women, as well as a sub-sample of 625 women which data were used in all but the simplest LCA model (i.e., in Model 1, all 920 women were involved, while in Models 2 through to 4, only 625 were included).

The reduction in cases from 920 to 625 deserves explanation here. In all but Model 1 (i.e., in Models 2, 3, and 4), the SPIn-W Pre-screen risk scale was utilized as a covariate in the latent class models, meaning that it was allowed to influence classification, probabilities, and overall model fit.

**Table 22**

*Sample Characteristics: Means (SD), Frequencies, and (Sub)sample Sizes*

Sample Characteristic	<i>M (SD) or n (%)</i>	
	Full Sample ( <i>N</i> = 920)	Subsample ( <i>n</i> = 625)
Age [Range]	34.52 (8.67) [20-70]	34.14 (8.18) [20-63]
Race (%)		
White	821 (89.2)	558 (89.3)
Asian	4 (0.4)	4 (0.6)
Black	26 (2.8)	17 (2.7)
Indigenous	53 (5.8)	38 (6.1)
Hawaiian/Pacific Islander	1 (0.1)	1 (0.2)
Mixed Race	10 (1.1)	3 (0.5)
Unknown Race	5 (0.5)	4 (0.6)
Index Offense (%)	<sup>a</sup>	<sup>b</sup>
Non-violent (non-drug)	434 (47.2)	316 (50.6)
Drug offenses	339 (36.8)	224 (35.8)
Violent (non-sexual)	128 (13.9)	79 (12.6)
Sexual	14 (1.5)	4 (0.6)
SPIn-W Pre-Screen Risk ( <i>M, SD</i> )		
Total risk	-	56.87 (18.22)
Dynamic risk	-	33.35 (13.42)
Static risk	-	23.52 (9.22)
Risk classifications (%) <sup>c</sup>		
Low	-	91 (14.6)
Moderate	-	506 (81.0)
High	-	28 (4.5)
General Recidivism rates (%)		
24-months follow-up	89/677 (9.7)	61/478 (12.76)
36-months follow-up	131/551 (14.2)	91/380 (23.95)
48 months follow-up	145/478 (15.8)	102/316 (32.28)
60 months follow-up	161/400 (17.5)	116/269 (43.12)

*Note.* “-” = data not available. General recidivism included any new offenses and technical violations that resulted in a return to custody to a Maine State Prison. <sup>a</sup>919 valid cases. <sup>b</sup>623 valid cases. <sup>c</sup>Risk classifications were derived using guidelines presented in Jones et al.’s (2015) publication; 0 (*no*), 1-37 (*low*), 38-88 (*moderate*), and 89+ (*high*).

The Pre-Screen measure is meant to function as a triage tool, such that only women with moderate and high risk on this brief version of the SPIn-W will go on to receive the full assessment. Thus, the expectation would normally be that all 920 women with full assessments also have valid Pre-Screen scores. However, during pre-analyses data verification and cleaning, it was discovered that the existing Pre-screen scores calculated by Orbis Partners, Inc were derived by treating both missing item-level data and scores of zero (0) as though they were definite indicators of women's absence of risk on any given item. This means that a woman with missing data on all items of a given domain would receive the same risk score as a woman who scored 0 (zero; no risk) on the same items.

The current author chose not to rely on the pre-calculated Pre-Screen risk scores that had been derived utilizing the method described above. If a woman with missing data on many items actually were at high risk in the areas covered by the items, the Pre-screen data would be inaccurate at best, and potentially invalid. Due to their threat to the validity of the Pre-Screen total scores, and thus any conclusion that could be drawn in analyses involving these scores, the current author used SPSS syntax to compute new Pre-screen risk total scores only for women with valid data on all items. Unfortunately, this led to a loss of 294 women, leaving a sample of 625. Both the original sample of 920 women, and the reduced 625 subsample were used in the current study, though the focus on the LCA remains on the 625 women with Pre-screen scores. It is made clear throughout the current study when, where, and for what purpose each of the samples were utilized.

Orbis Partners, Inc: A brief note. Orbis Partners, Inc is a private corporation providing risk/need/case management assessment tools for a variety of correctional populations- including mixed gender adult (the Service Planning Instrument [SPIn]; Orbis, 2003) and youth samples

(Youth Assessment Screening Instrument [YASI]; 2000), as well as justice-involved women specifically (SPIn-W; Orbis Partners, 2006) across North America. The instruments are used both in custodial and community settings. Despite severely limited validation research on some tools-including the SPIn-W- existing findings show some promise in terms of the various tools' ability to predict recidivism with adequate accuracy. The Measures section below describes the findings of validation studies of the SPIn-W.

## **Measures**

### ***The Service Planning Instrument for Women (SPIn-W; Orbis Partners, 2006)***

The SPIn-W is a risk and needs assessment and case management tool developed specifically for adult women in both institutional and community settings. The SPIn-W is a version of the original gender-neutral SPIn Instrument, which was widely used in North-American jurisdictions to assess both men and women (Orbis Partners, 2003). The original SPIn itself was based on the Case Management Assessment Protocol (CMAP; Barnoski, 2003), a well-known instrument originally utilized with juvenile justice clients in Washington State. The SPIn-W is based on the RNR framework, with the well-established, proposed gender-neutral central eight risk/need factors still included in the assessment (Andrews, 2012). However, this tool also encompasses additional factors proposed by gender-responsive scholars to be relevant to women's recidivism risk and the identification of treatment targets (Chesney-Lind & Shelden, 2003; Van Voorhis et al., 2010). Some examples of these new additional items include but are not limited to self-efficacy, emotional regulation, familial and romantic relationships, child rearing, and mental health.

Assessing individual women using the SPIn-W is a step – by – step process starting with the completion of a semi-structured interview, followed by a review of all available collateral

sources (e.g., criminal records, case files, treatment reports, mental health evaluations, employment records, and data gleaned from interviews with family and friends). In total, the SPIn-W consists of 100 main items and a very large number of sub-items, divided across 11 domains: Domain A – Criminal History (six items), Domain B – Response to Supervision (10 items), Domain C – Family and Children (12 items), Domain D – Social Network (seven items), Domain E – Substance Abuse (five items), Domain F – Employment (eight items), Domain G – Attitudes (six items), Domain H – Social/Cognitive Skills (12 items), Domain I – Mental Health (seven items), Domain J – Violence (16 items), and finally, Domain K – Community Living (eight items).

### **Item Response Scales and Scoring**

The response scale and measurement of the 100 items and their sub-items vary. Some items take a binary *yes vs. no* format (e.g., “Physical abuse as a child”), some are measured continuously (e.g., age), and many are measured on a 5-point Likert-type scale. For example, response options for Item 5 of the Employment Domain “Job Search Skills” range from 0 (*Poorly informed of labor market opportunities*) to 4 (*Confident about job search techniques*). Although only risk scores are utilized in the current study, it is worth mentioning that many of these Likert-type items receive both risk and protective scores, separately. For example, using the item above, a score of 1 (*Not confident about ability to find employment*) would receive a risk score of 1 and a protective score of 0. Risk and protective scores do not influence one another; the final risk score remains the same whether the woman scores high or low on protective factors.

It is important to note that typically, the Pre-Screen assessment is administered first to triage clients; low risk cases are not administered the full assessment. Only individuals found to

be moderate or high risk to reoffend subsequently receive the full SPIn-W assessment. A short description of how scores are derived for both the Pre-Screen and the full assessment is provided below.

Assessors enter individual item responses into the SPIn-W software program, which sums and converts individual responses into a series of static, dynamic, and overall total risk ratings<sup>19</sup> based on the Nuffield weighting scheme (Nuffield, 1982). Both Pre-screen and full assessments are derived using this technique. For both assessment versions, the program returns a risk report, a risk classification (i.e., *low, moderate, or high risk*), and a protective (i.e., strength) classification (i.e., *low, moderate, high*). When the full assessment is employed, the program also generates a circular graphic called the SPIn Wheel, the latter of which used as a visual aid to the assessor in identifying the key areas of concern for the individual woman (see Figure 5 below). The wheel is not produced for the Pre-screen measure (personal SPIn-W assessment training, October, 2020). Notably, classification norms can, and are frequently adjusted to suit the needs of individual jurisdictions (Jones & Robinson, 2018).

### ***Mental Health Indicators: Items Derived from the Full Assessment***

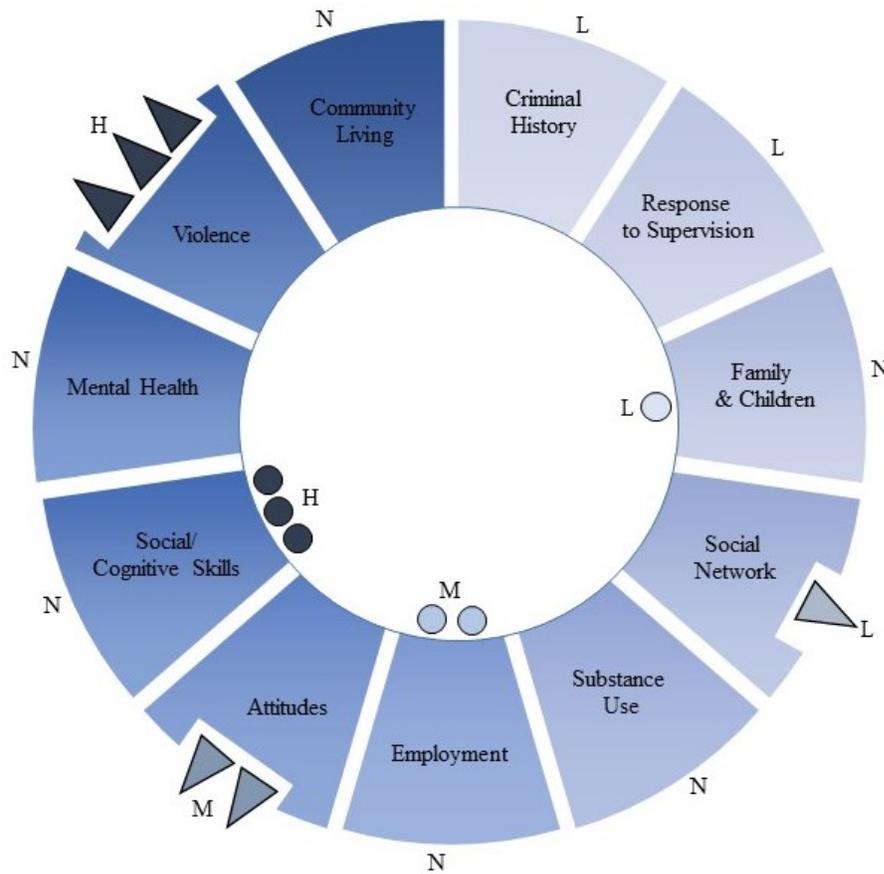
Domain I of the SPIn-W is focused on mental health and contains seven main items, some of which are associated with a number of smaller sub-items. Appendix O lists all the mental health items included in Domain I that were utilized as indicators in the LCA; 14 in total. There were nine clinical disorders: Depression or other affective disorder (mood disorder), anxiety disorder, bipolar disorder, BPD, thought and adjustment Disorders, PTSD, schizophrenia, other psychoses, and ADHD.

### **Figure 5**

---

<sup>19</sup> Static, dynamic, and total protective scores are also derived.

*The SPIn-W Wheel (reconstruction; Jones & Robinson, 2018)*



*Note.* The Spin-W Wheel presents risk and protective classifications on each domain of the full assessment. The main blue pie chart are presents the individual domains. Tringles indicate level of risk with one representing low risk, two representing moderate risk, and three representing high risk. The small circles nested within the chart indicate level of protective factors on each relevant domain, with one representing low, two representing moderate, and high representing high protective scores. L = low. M = Moderate. H = High.

. Given the similarity in symptomology and low prevalence rates, schizophrenia and psychoses were combined to a single category. Only currently active disorders were included. Items speaking to sexual, physical, and emotional abuse throughout the lifespan were not included as indicators in the LCA, but were included in a measure of Adverse Childhood Events (ACEs; see description of this measure in the next section below), which total score was used as an auxiliary variable in the LCAs. Further, binary items assessing eating disorders (Item 3a),

self-injurious behaviour (Item 3b), somatization (Item 3c), and homicidal ideation (Item 4) were included as individual indicators. Lastly, Item 5 was originally measured on a 3-point Likert scale from 0 (*no indications*), 1 (*suicidal thoughts*), and 2 (*suicide attempts*). For the purposes of the current LCA, this item was dichotomized into *no indications* vs. *suicidal thoughts and/or ideation*. An additional item was included as an indicator in LCA “Comorbidity”, which simply measured the co-occurrence of more than one mental disorder.

It is worth noting, that as part of additional exploratory post-hoc investigations of class differences in terms of mental health, two additional items were computed including “*Co-Occurring*”, which response options provided the number of women who had *two to three* and *four to six* different mental disorder currently (i.e., a measure of complexity), and “*Any Mental Disorder*”, which simply provided the frequency of women with at least one diagnosed current disorder.

***SPIn-W: Pre-Screen (see Appendix J for copy).*** The SPIn-W Pre-Screen is a 35-item screening tool used for triage and classification purposes (Jones & Robinson, 2018) for adult women; higher scores = higher risk. It includes a subsection of items from all but one domain – Domain K (DK) –Community Living. Specifically, it consists of six Criminal History (DA), six Response to Supervision (DB), three Family and Children (DC), two Social Network (DD), one Substance Abuse (DE), three Employment (DF), two Attitudes (DG), two Social/Cognitive Skills (DH), five Mental Health (DI), and finally five Violence (DJ) items. Appendix J contains all the items on the SPIn-W Pre-Screen by domain. Note that the Pre-Screen scale is a triage tool meant to identify moderate and high-risk women who are believed to need a full assessment. Lower risk women are typically screened out and do not receive the full assessments. Thus, most women with full assessment scores will necessarily have a moderate to high estimated level of risk. In

the current study total raw score was used rather than the 4-level categorical risk classification scheme.

The validation of the SPIn-W is still in its infancy<sup>20</sup>. Very little empirical research that could speak to its reliability, validity, and predictive utility has been done to date. Indeed, even for the original gender-neutral SPIn measure (Orbis, 2003)-the precursor to the SPIn-W- that has been in use for a much longer period of time than it's women-centered descendant, there are only four studies that have investigated the predictive validity and these focused on the Pre-Screen version (Jones et al., 2015, Jones & Robinson, 2017a, 2017b, 2018, as cited by Goodwin, 2020). AUC values have been found to range from a very modest .59 to a very favorable .77 (Rice & Harris, 2005). AUC values are straightforwardly interpreted; An AUC = 50 indicates prediction at chance level: An AUC of 1 indicates perfect prediction. Jones and Robinson (2018) also investigated the reliability of the pre-screen SPIn-W using the sample of 274 adult women probationers in Connecticut. Internal consistency for individual domains was *good* or *very good* for six domains ( $\alpha = .75$  to  $.82$ ), poor for two ( $\alpha = .60$  to  $.65$ ), and very poor for the remaining three domains ( $\alpha = .39$  to  $.54$ ). It's important to remember that Cronbach's alpha is heavily influenced by the number of items included in its scale reliability estimation (Schmitt, 1966).

### ***Adverse Childhood Experiences Scale (ACEs)***

A SPIn-W-based Proxy Measure of the ACEs was constructed for the current study (Felitti et al., 1998; Wanamaker, 2020). The proxy measure of the ACEs used in the current study is available in Appendix K below as well as in the already included in full in text Table 23.

---

<sup>20</sup> Note that the instrument's predictive validity also depends on the intersection between the demographics of the target population and the weighting method (i.e., algorithm) used (e.g., no weighting, The Burgess Method, the Nuffield Method; see Robb, 2020).

Adverse Childhood Experiences (ACE) are a diverse set of traumatic experiences during childhood (ages 0 to 18 years of age). The ACE scale (Felitti et al., 1998) was designed to assess such experiences using a selection of 10 binary (*yes* vs. *no*) questions, summed to a total score ranging from 0 (zero) to 10. The ACES has become the most commonly used measure of adversity (i.e., potential trauma) in childhood, and has been found relevant to both the development of serious mental health disorders in adulthood (Messina & Grella, 2005), as well as criminal behaviour (Messina et al., 2007), as discussed in detail in the Introduction section above. Orbis incorporated the original 10 ACES items, along with a few other related indices of childhood adversity as part of the SPIn-W measure, in order to improve the assessment of this proposed gender-responsive risk factor (i.e., trauma). In the current study, a proxy measure of the ACES was utilized. Its basis was a well-validated and widely used early revised version of the scale defined and utilized by the Centers for Disease Control and Prevention (n.d.). This variant of the ACEs includes four main domains; Childhood Abuse (e.g., “Physical Abuse”), Childhood Neglect (e.g., “Emotional Neglect”), Household Dysfunction (e.g., “Mental Disorder in the household”), and finally Unstable Home (e.g., “Parental Divorce/separation”). All of the 10 standard items were available from the SPIn-W. In addition, two other items were included to help better inform childhood adversity. Nevertheless, the total scores could still range from 0 to 10 (i.e., additional items only counted toward the maximum possible score in their own existing domain), with the lowest score indicating no exposure and the highest score indicating the most extreme exposure. Appendix K contains all items divided across the four domains just presented above.

Previous studies have reported favorable psychometric properties of the original ACEs and its many, varied proxy derivatives. For example, Wanamaker (2020) utilized the gender-

neutral original SPIn measure to create an ACEs proxy based on eight items. This proxy measure evidenced adequate internal consistency (Cronbach's alpha for women;  $\alpha = .71$ ). This method has also been validated in previous studies (Baglivio et al., 2015).

In the current study, internal consistency for the total scale was moderate with  $\alpha = .72$  ( $n = 625$ ). Of note, reliability analyses included the two additional items because results of assessing potential internal consistency improvements to their exclusion indicated that removing them would degrade overall internal consistency.

### ***Recidivism***

Recidivism information was obtained from official records collected by Orbis Partners from the State of Maine. Recidivism here was defined any new offense or any technical conditional release violations that resulted to a return to custody in a Maine state prison. It is important to understand that recidivism that resulted in detainment in a local remand jail, or in a state prison outside of Maine were not registered as recidivism for this sample. This restricted recidivism measure is the underlying reason for the problematically low the base rates for all follow-up times in the sample. Table 22 above containing the sample characteristics demonstrate the very low base rates relative to the averages reported for large, comparable sections of the justice-involved population of adult women across the United States (Deschenes et al., 2007)

Correctional researchers have found that returns to custody is the best recidivism outcome measure in studies of adult women (Lynn-Stewart et al., 2019). The length of follow-up selected as the main measure of recidivism in the current study was 24 months. This was selected to minimize sample size loss and is consistently used in the recidivism literature (Lynn Stewart et al., 2019) had the largest proportion of women with valid data. Thus, it was selected in order to avoid unreliable results on account to unacceptably low base rates ( $< 10\%$ ; Tabachnick &

Fidell, 2019; See Table 22 above). The number of women with available follow-up data varied depending on the length of follow-up time, and base rates naturally increased with increases in follow-up.

### **Procedure & Acquisition of Data**

The archival dataset utilized in the current study was provided by Orbis Partners, Inc, located in Ottawa, Ontario, Canada. Orbis is a Canadian correctional consulting firm that provides risk/need/case management assessment tools for a wide variety of specific correctional populations and jurisdictions, develops correctional programs, and provides training to a variety of agencies and individuals working with justice-involved populations.

A Memorandum of Understanding (MOU) between Orbis, the Maine Department of Corrections, Carleton University, the project supervisor Dr. Shelley Brown, and finally, the principal researcher and current author-Cathrine Pettersen- was signed, granting permission to use of the archival dataset (Appendix M). Ethics clearance was obtained from Carleton University's Ethics Board and is presented in Appendix N.

### **Analytical Approach: Preparatory & Exploratory Analyses**

The Statistical Software Package for the Social Sciences (SPSS) was used for data screening, verification (e.g., identifying and correcting any data errors<sup>21</sup>), variable computation, and missing data management. Latent class analyses were conducted Mplus-a statistical software program designed exclusively for latent variable modeling. LCA procedures and analyses are described directly following the current preparatory general analyses and procedures in the section directly following the current description of preliminary analyses.

---

<sup>21</sup> An example of a data error might be the erroneous exclusion of one subscale in the calculation of a total scale score, leading to incorrect total scores.

### Data Screening

All variables were examined for presence of any data errors, missing variables that had to be computed to facilitate analyses, missing data, and violations of statistical assumptions relevant to the current analyses. In terms of missing data on the LCA mental health indicators, there were few missing data.

Before progressing to the mental health indicators, it is worth noting that age was available for all 920 women and that only a single woman did not have an index offense listed. However, of the entire sample, 26.4% ( $n = 243$ ) were lost during the shortest follow-up period of 24 months, such that recidivism rates in Table 22 pertain to only 677 women. Longer follow-up periods resulted in greater loss of data (see Table 22).

Of the initial sample of 920 women, a very small percentage (2.28%) had missing data on Item 4 (“Homicidal Ideation”). For the remaining mental health indicators ( $n = 1$ ), the additional mental health variables used exclusively in post-hoc analyses ( $n = 1$ ), as well as the ACEs ( $n = 1$ ), there was only one missing value – a single woman had missing data on most variables. Because the relationships between Homicidal Ideation and the remaining variables were weak, the number of missing data points small, and analyses did not disprove the assumption that data are missing completely at random (MCAR), no action was taken to eliminate missingness on this variable. The single case with missing data on almost all variables of interest was naturally excluded as part of standard procedures in Mplus.

**Table 23**

*LCA Indicators & Additional Mental Health Variables: Sample Proportions (N = 920; n = 625)*

Mental Health Indicators & Related Variables	<i>n</i> (endorsed item)/total <i>N</i> (%)	
	Total Sample ( <i>N</i> = 920)	Subsample ( <i>n</i> = 625)
<b>Current Mental Disorder Diagnoses</b>		
Mood disorders	310/919 (33.7)	213 (34.1)
Bipolar disorder	158/919 (17.2)	114 (18.2)
Anxiety disorders	464/919 (50.4)	317 (50.7)
Thought/adjustment Disorders (Any)	28/919 (3.0)	21 (3.4)
Borderline personality disorder (BPD)	32/919 (3.5)	22 (3.5)
Post-traumatic stress disorder (PTSD)	200/919 (21.7)	145 (23.2)
Psychotic disorders (Any)	26/919 (2.8)	20 (3.2)
Attention-deficit/hyperactivity disorder (ADHD)	59/920 (6.4)	39 (6.2)
.....		
Any mental disorder (current)	611/919 (66.4)	417 (66.7)
Including eating disorders	637/919 (69.2)	439 (70.2)
Comorbid mental disorder	275/919 (29.9)	248 (39.7)
Co-occurring disorders		
2-3 disorders	-	185 (29.6%)
4-6 disorders	-	63 (10.1%)
<b>Behavioural Indicators</b>		
Somatization	2/919 (0.2)	2 (0.3)
Eating disorders (Any)	120/919 (13.0)	88 (14.1)
Self-harm (past/current)	192/919 (20.9)	138 (22.1)
Suicidality (ideation/attempts; past/current)	189/919 (20.5)	133 (21.3)
Past suicide attempts	-	111 (17.8%)
Homicidal ideation (current)	23/899 (2.5)	15 (2.4)

*Note.* “ - “ = these variables were added in post-hoc analyses and proportions were thus limited to the LCA subsample; they were not included as LCA indicators but are rather reported here to provide a better overview of the severity and complexity of mental health problems in the sample.

Of note, the original dataset contained Pre-screen total risk scores for a majority of the overall sample. However, it was discovered that in order to calculate these scores, missing data (i.e., item responses denoted -1) had been automatically considered the same as scores of 0 on the same items. This means that women for whom data were not available on a given item received the same risk score as women who had received a risk score of 0 (zero) on the same item. Given the potential for such scoring techniques to obscure important differences in risk between individual women, in particular if women with missing data differed systematically on risk relative to women with valid data on all items, alternative Pre-Screen total risk scores were calculated separately for the current analyses. A Pre-screen total risk score was only calculated for women with valid data on all Pre-screen items. As a result, 295 had missing total risk scores on the Pre-screen measure. This loss of data did not impact the first LCA model, as this model was based only on the 14 individual mental health indicators. However, for the remaining models- Models 2, 3, and 4, the sample size included in analyses was reduced to 625, on account of the Pre-screen total score being included in the models as a covariate.

### **Assumption of Distributional Normality**

The distributions of all count and continuous variables included in any part of the current study, were examined in terms of normality, skew, and kurtosis. ACEs scores were somewhat negatively skewed with the vast majority of women scoring between 0 (zero) and six (out of 10). The Pre-screen total score to be included in LCA in Models 2 through 4, was normally distributed, and did not evidence any problematic skew or kurtosis. In conclusion, no data manipulation or transformations aimed at altering the distributions of variables were employed in this study.

### **LCA Procedures: Preparations for a Multi-Stage LCA Analysis**

This section describes the preparatory procedures and the LCA approach. First, testing of the basic statistical assumptions of latent class analysis is briefly described. Next, information regarding software, syntax construction and key analysis specifications (e.g., specification of mixture modeling) are provided in a short summary. Further, the logical sequence in which the main analyses were conducted are presented to provide the foundation for an understanding of the final results and to ease interpretation. Lastly, the crucial elements involved in model evaluation are presented.

### **LCA Assumption Testing**

As part of initial preparatory analyses conducted in SPSS, data were further tested in terms of violations of assumptions pertaining primarily to LCA modeling. Chi-Square analyses that are used in the LCA procedures require minimum expected cell frequencies of five. A single indicator- "*somatization*"- violated this assumption, which was endorsed by only two women. Despite this violation, this was the only indicator without adequate cell counts; it was retained in analyses to maximize the data utilization. It is worth noting that LCA models were also tested excluding "*somatization*" and two other rare indicators; "*homicidal ideation*" and "*thought and adjustment Disorders*". Appendix L summarizes the findings for models excluding these three items. Note here that the decision following testing was to include them as intended in the main models (see Appendix L).

Several additional indicators initially violated the assumption of minimum proportions of 10% per cell. "*Borderline personality disorder*" (BPD), "*psychotic disorders*", "*ADHD*", and as mentioned, "*thought and adjustment Disorders*" and "*homicidal ideation*" violated this assumption. In the full and LCA subsamples ( $N = 920$ ;  $n = 625$ ), their proportions ranged from

about 2.5% to 6.5% (see Table 23). These indicators too were retained however, as it was discovered that although they had percentage representations considered problematic in the initial samples, when utilized in LCA models with two- and three-class solutions, all but one (“Homicidal Ideation”) had proportions just shy of 10% or considerably higher within a single class. Thus, these indicators added unique information to the class structure and promoted the interpretation of the classes, showing non-negligible, meaningful variability across classes.

Finally, modeling latent classes requires the indicators to be locally independent. Local independence specifies that within one latent variable (i.e., class), the indicators are independent. Put differently, if it had been possible to create separate contingency tables the indicators that corresponded to each latent class, the indicators within these tables would be independent. However, the latent class cannot be observed directly, and thus the contingency tables can only be approximated (Collins & Lanza, 2010. Mplus conducts this approximation and reports its results in standardized output automatically: for all main models, Mplus output confirmed local independence.

The statistical software program designed for latent variable modeling Mplus 8.1 was used to conduct the latent class analyses. Syntax was used to specify the Multiple Likelihood Ratio (MLR) as the estimator and the analysis type as Mixture. The integration algorithm was utilized (Muthèn & Muthèn, 1998-2017). For the main models reported in detail the LCA Results section below, random starts were specified as 1,000 and 250. Model covariance coverage was set to a minimum of 0.100 as recommended by Muthèn and Muthèn (1998-2017). Research questions regarding appropriate latent structure was examined in successive LCAs in a stepwise fashion. The purpose of sequentially testing Models 1 through 4, adding only one additional

variable at a time was to further elucidate the contributions of each variable to overall model fit and the nature and subsequently the interpretation of classes.

The least complex model was tested first; Model 1 included only the 14 mental health indicators and fit was estimated for a single class, as well as two, three, and four classes. Indicators are those variables that are utilized directly in the analysis model to inform latent class formation. Second, Model 2 added the Spin-W Pre-Screen raw total score as a model covariate, again examining the accuracy with which the data was reflected when a latent class structure specifying one through to four classes was imposed. Briefly, a covariate is a variable that can be understood to be a predictive, causal factor of the latent classes. A covariate can improve classification accuracy and given that it is thought to impact class membership, it is necessarily controlled for during the construction of the classes (Lubke & Muthèn, 2007). Third, Model 3 added three auxiliary variables, ACEs, race, and age. Briefly, auxiliary variables, unlike covariates, are not included directly in the analysis model, but are variables added after the initial LCAs to provide additional insights into the distribution non-indicator characteristics across latent classes. Auxiliary variables do not impact classification accuracy or class membership (Lubke & Muthèn, 2007). Finally, Model 4- the most complex model and the one ultimately of primary interest in the current study- incorporated all the elements of the previous models, but further added the distal outcome of recidivism at 24 months follow-up. A distal outcome can be understood as a consequence of class membership. Following initial analyses of the four models with the four latent class solutions, post-hoc analyses involving pairwise comparison of classes were also conducted to further clarify the nature of the derived classes.

### **Evaluation of Model Fit and Latent Class Solutions**

Contemporary methodologists (e.g., Marsch et al., 2009) recommend relying on multiple measures of the fit of various class structures, including both statistical indices and other, more subjective methods of interpretation. In line with these recommendations, to evaluate the fit of the proposed class solutions, a combination of statistical fit indices assessing overall model fit, Likelihood Ratio Tests speaking to the viability of the  $k$  model (current model) versus the  $k-1$  model (i.e., the class solution with one less class than the one currently under investigation), local fit in terms of patterns of mental health indicators, classification accuracy, class proportions, and the meaningfulness and interpretability of classes were employed to determine the viability of each model and class solution. Overall model fit was evaluated using a number of criteria, one of which was the evaluation of global fit indices; the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Sample-size Adjusted Bayesian Information Criterion (SSA-BIC). The AIC is a measure of the distance between the specified and the true model, while the BIC and SSA-BIC measure the relative accuracy of the predicted probability distribution and the probability distribution proposed by the specified model (Vrieze, 2021). There are no absolute cutoff scores applicable to these fit indices. Rather, when various class solutions are compared the model generating the lowest absolute values was preferred (Muthén & Muthén, 1998-2017). Entropy was also considered; Entropy is a measure of classification accuracy (i.e., it is a measure of the degree of disorder in the model with the specific number of classes specified). Values  $\geq .80$  are considered indicative of adequate accuracy (Muthén & Muthén, 1998-2017).

Further, classification accuracy (posterior probabilities) were examined directly; values  $\geq .70$  are considered adequate, as well as class proportions- whether or not the identified additional class represented a meaningful portion of the sample in a conceptually meaningful way.

Next, by examining the probability scale indicating a given class members' probability of presenting with each mental health indicator, the utility of each additional class could further be evaluated in terms of what additional information the added class(es) could provide about the underlying structure of the data. Lastly, the likelihood ratio tests (LRTs) – the Vo-Loung-Mendell-Rubin LRT (VLMR-LRT) and the Lo-Mendell-Rubin LRT (LMR-LRT) – that directly assess model improvement for the current model ( $k$ ) versus a model with one less class ( $k-1$ ), were then consulted. The smaller the probability value of the resulting LRTs, the clearer the support for the  $k$  model – the model with the added class- over the previously tested  $k-1$  model.

## Results

### Preliminary Analyses: Pre-LCA Explorations of Mental Health & Adversity

Following these initial preparatory procedures, the SPSS dataset was used to obtain descriptive statistics (e.g., means and standard deviations [ $SD$ ]), frequencies and proportions for all descriptive and study variables for the full sample ( $N = 920$ ) and the reduced subsample for LCA Models 2, 3, and 4 ( $n = 625$ ). Refer to Table 23 above below for frequencies and proportions of women endorsing the 14 mental health indicators.

I hypothesized that between 70% and 80% of women in the sample would meet diagnostic criteria for at least one current mental disorder. As is clear in Table 23 above, the rate of women with any (i.e., at least one) mental disorder fell marginally short of these expected rates with roughly 66% to 70% endorsing at least one disorder. Anxiety disorders were among the most common, with approximately half of all women having at least one current diagnosis of

an anxiety disorder. Mood disorders such as depression were also common, affecting about 30% of both the full and the reduced sub-sample. The rates of PTSD were also high, hovering around 20%. Surprisingly, rates of borderline personality disorder (BPD) were much lower in this sample than in samples assumed to be comparable a-priori (i.e., 3.5% vs. more than 30% among federally incarcerated women in Canada; OCI, 2019)

Further, I hypothesized that rates of non-lethal self-injury, suicidal ideation, and suicide attempts would be a recognized problem for 10% to 25% of the overall sample. In line with expectations, self-harm and suicide affected approximately 20% of both the full and the reduced subsamples, supporting a-priori hypotheses based on findings in previous studies.

Next, I tentatively hypothesized that 15% to 20% of the sample could be classified as severely mentally ill. As mentioned above, the rate of severe mental health needs is difficult to determine because this construct lacks any formal definition as well as any consistent measurement. Recall for example from Study 1 in the current program of study – the meta-analysis examining the aggregate relationship between mental disorders and recidivism- different authors' definitions of what constitutes a severe mental disorder varied rarely overlapped between studies. Nevertheless, some indications were present in the current sample. First, 30% to 40% of all women in the full and reduced subsamples, respectively, suffered current comorbid disorders. One-hundred and eighty-five women (29.6%;  $n = 625$ ) suffered two to three concurrent disorders, while 63 (10%) suffered between four and six currently active co-occurring disorders. Moreover, 17.76% reported having attempted suicide in the past. Taken together, the obtained proportions appear to support the hypothesized estimate of at least 20% suffering severe mental health challenges.

## ACEs

Further, the rates of exposure to adverse childhood experiences (ACEs) were also examined for the current samples ( $N = 920$ ;  $n = 625$ ). Table 24 below presents the mean ACEs scores ( $SD$ ) as well as frequencies of exposure to the 12 individual adversities. On average, women in the current sample had experienced between three and four adverse childhood events (i.e., potential traumas) at some time between the ages of 0 and 18, which is somewhat lower than the average reported in other recent studies of justice-involved women (e.g., Vaswani, 2019;  $M = 6.07$ ,  $SD = 2.7$ ). The most frequent type of trauma reported was childhood abuse. Rates of physical, sexual, and emotional abuse fell in the range of about 40% to 60%, with sexual abuse being the most common. These rates are comparable to those reported in previous studies of justice-involved women (e.g., Robinson et al., 2012). Childhood neglect was reported by between 6% and 10% of women.

Approximately half of the sample reported that substance abuse in the family contributed to household dysfunction, approximately 10% to 14% reported either mental illness or the incarceration of a family member during childhood, and as much as 25% reported witnessing domestic violence against their mother growing up as well as parental divorce and separation.

Around 35% reported that their childhood environment was characterized by frequent conflicts with the parent(s) and just over 16% reported being placed in foster care or equivalent at least once.

**Table 24**

*Adverse Childhood Experiences (ACEs): Means (SD) & Endorsement Frequencies (%)*

ACEs <sup>a</sup> Scores & Item Frequencies	<i>M (SD) or n (%)</i>	
	Total Sample ( <i>n</i> = 919) <sup>b</sup>	Subsample ( <i>n</i> = 625)
ACEs Total Score	4.28 (2.48)	3.30 (2.42)
Childhood abuse		
Childhood sexual abuse	536 (58.4)	384 (61.4)
Childhood physical abuse	386 (42.0)	271 (43.4)
Childhood emotional abuse	426 (46.3)	303 (48.5)
Childhood neglect		
Childhood physical neglect	59 (6.4)	38 (6.1)
Childhood emotional neglect	100 (10.9)	71 (11.4)
Household Dysfunction		
Household substance abuse	459 (49.9)	319 (51.0)
Household mental illness	131 (14.2)	89 (14.2)
Household member incarcerated	100 (10.9)	58 (9.3)
Mother treated violently	232 (25.2)	171 (27.5)
Unstable home		
Parental divorce/separation	227 (24.7)	153 (24.5)
Foster placements	151 (16.4)	102 (16.3)
Frequent conflict with parents	327 (35.5)	232 (37.1)

*Note.* <sup>a</sup>ACEs = Adverse Childhood Experiences Scale (proxy). <sup>b</sup>One woman had missing data on ACEs.

## LCA Results

LCA modeling was conducted in Mplus 8.1. The above analyses and procedures section outlines syntax specifications and basic procedures employed with this software program. Appendix P contains the full Mplus syntax for the final full model (i.e., Model 4 with a three-class solution, the Pre-Screen total risk score covariate, ACEs, age, and race as auxiliary variables), and 24-month recidivism as the distal recidivism). For all but two models and solutions- Models 3 and 4 with four-class solutions- the best loglikelihood value was successfully replicated and model estimation terminated normally. The footnote below describes the problem of model non-convergence for these models and solutions and tests conducted to confirm that this issue was due to gross model misspecification (i.e., that the four-class solution simply did not account for the latent structure of the observed data rather than, for example, syntax specification or data errors)<sup>22</sup>.

The sections below are organized sequentially. Findings pertaining to the initial Model 1 containing only the 14 mental health indicators are presented first. Note that following the basic in-text description of Model 1 findings, the main LCA summary Table 25 containing all statistical indices used to evaluate overall model fit across all models and solutions is presented.

---

<sup>22</sup> Methodologists such as Muthén & Muthén (1998-2017) have outlined a number of syntax corrections and additions that can be employed to resolve non-convergence problems that are not due to model misspecification but rather to problems with the data or the technical specifications made. These include varying the number of random starts, adding specifications of STITERATIONS (i.e., the maximum number of iterations allowed in the initial stage of random starts; Muthén & Muthén, 1998-2012), and disabling random starts in favour of researcher-specified individual indicator thresholds. All three methods were fully tested. The persistence of model non-convergence confirmed that gross model misspecification was the only plausible cause of these errors.

A return to this summary Table 25 is then recommended in order to view fit indices of interest for Models 2 through to 4. Second, Model 2, in which the SPIn-W Pre-screen risk score was added as a covariate is described. Third, Model 3, ACEs, age, and race were added as auxiliary variables. Lastly, Model 4 is presented; this model retained all the elements included in the three previous models with the addition of 24-month recidivism as a distal outcome. Note that for each model, the class solution (one, two, three, or four) that best represented the data, statistical indices of global model fit (consult Table 25), classification accuracy, class proportions, and a brief, evolving description of the resulting classes are reported. Full class descriptions and final class titles are subsequently presented in a final, separate section. Premature naming and detailed descriptions of classes prior to the final evaluation of the complete model (i.e., Model 4) would likely produce misleading, inaccurate, and possibly biased interpretations and descriptions. Thus, this content is presented last.

### **Model 1: What the 14 Mental Health Indicators Revealed**

Model 1 involved only the 14 mental health indicators and was examined with one-, two-, three-, and four-class solutions. The model fit indices, entropy, and the VLMR and the LMR likelihood ratio tests (LRT) are presented in summary Table 25 below for the current Model 1 as well as the remaining Models 2 through 4 with one- through four-class solutions. A two-class solution provided superior fit relative to the single-class structure. The two resulting classes differed in terms of the proportions of women endorsing single indicators as well as the number of indicators endorsed; One class had severe mental health problems, the other did not.

Despite favourable LRT values, acceptable entropy, perfect classification accuracy (i.e., latent class assignment and most likely latent class membership [posterior] probabilities corresponded perfectly), the two-class solution was ultimately rejected.

Including a third class improved model fit substantially (refer to summary Table 25 above). Classification accuracy for the three-class solution was reduced minimally relative to the simpler two-class structure, as expected with any increase in solution complexity, but remained high and well above the minimum recommended threshold of  $\geq .70$ .

Inclusion of this third class revealed crucial differences in the mental health of the women in this sample. In the three-class solution, Class 1 – temporarily referred to as “Severe Mental Health Needs”, remained intact. However, a very large portion of Class 2- originally labelled “moderate mental health needs”- was separated into Class 3. The novel Class 3 had very low rates of overall mental health needs and was termed “Low Mental Health Needs”. The nature of the remaining Class 2, however, was of most interest- it was here that the novel finding made itself known. In this class, mental disorders did occur with some frequency, but at a much-reduced rate relative to the severe needs Class 1. However, despite these more modest prevalence rates overall, Class 2 now had alarmingly high rates of self-harm and suicidal thoughts and behaviours- even as compared to the most severely troubled Class 1. About 30% of women in Class 1 endorsed self-harm and suicide indicators, but in Class 2, these numbers increased to 42.2% and 51.2%, respectively.

**Table 25**

*Fit Indices, Entropy, & Likelihood Ratio Tests: All Models with One- to Four-Class Solutions*

Model	Variables included (N)	Classes	Indices of Overall Fit of Model and Class Solutions					
			AIC <sup>a</sup>	BIC <sup>b</sup>	SSA-BIC <sup>c</sup>	Entropy	VLMR <sup>d</sup>	LMR <sup>e</sup>
1	14 mental health indicators (N = 920)	1	9,439	9,506	9,462	-	-	-
		2	8,197	8,337	8,245	0.999	-4,705***	1,260***
		3	8,055	8,267	8,128	0.83	-4,070*	170*
		4	7,979	8,264	8,076	0.81	-3,984↓	105↓
2 & 3	14 mental health indicators SPIn-W Pre-Screen covariate Auxiliaries: ACEs, age, & race (n = 625)	1	14,854	14,931	14,880	-	-	-
		2	5,693	5,826	5,731	1.00	-3,270***	899***
		3	5,542	5,746	5,600	0.85	-2,817***	181***
		4 <sup>j</sup>	5,551	5,826	5,629	0.77	-2,725**	23**
4	14 mental health indicators SPIn-W Pre-Screen covariate Auxiliaries: ACEs, age, & race Distal outcome: recidivism (n = 625)	1	15,383	15,465	15,411	-	-	-
		2	6,062	6,204	6,103	1.00	-3,453***	900***
		3	5,855	6,077	5,918	0.85	-2,970***	184***
		4	5,918	6,211	6,001	0.87	-2,906	25

*Note.* All indices are reported to their closest whole integer for ease of read. Entropy values have been reduced to two decimals. The best fitting class solution in each model is highlighted **green**. Rejected/non-convergent models are highlighted **grey**. “-” = indices not available for single class solutions. Models 2 and 3 are reported in a single row because Model 3 was identical to Model 2 with the exception of three added auxiliary variables that do not impact any aspect of the models (i.e., fit indices are identical). <sup>a</sup>AIC = Akaike Information Criterion. <sup>b</sup>BIC = Bayesian Information Criterion. <sup>c</sup>SSA-BIC = Sample Size Adjusted Bayesian Information Criterion. <sup>d</sup>VLMR = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. <sup>e</sup>LMR = Lo-Mendell-Rubin Adjusted Likelihood Ratio Test. <sup>j</sup> = Model misspecification resulting in model non-convergence.  
 \*\*\*  $p < .001$ . \*\*  $p < .01$ . \*  $p < .05$ . ↓  $p < .10$

Class 2 was therefore temporarily renamed “moderate mental health needs with severe externalizing behaviours”. Despite the presence of Classes 1 and 2, both of which had severe mental health concerns, the largest class remained Class 3- the low mental health needs class- making up just over 54% of the sample.

A fourth and final class solution was tested next but non-significant LRT tests, meaning that a fourth class did not improve the model beyond the three-class solution. Note that for Model 1, the four-class structure was rejected absolutely using all available methods of model evaluation. Further, relevant output suggested systemic unreliability in terms of basic statistical coefficients (e.g., indicator probability scales relative to class proportions). As detailed in the individual model sections for Models 2, 3, and 4 below, with the exception of Model 1, attempts to test a four-class structure most often resulted in model non-convergence that was indisputably attributable to gross model misspecification. As a result, similar to outcomes for Model 1, values of several key statistical indices were discovered to be wholly unreliable. For this reason, statistics pertaining to four classes are not part of the main model results here or for the remaining models. The optimal solution was determined to be the three-class structure, which thus became the preferred class structure for Model 1. Table 26 directly below provides class proportions for all class solutions for the current Model 1 as well as the remaining three models.

### **Model 1 Summary**

Taken together, findings obtained with Model 1 can be summarized into four main take-home messages. First, a three-class solution best represented the observed data, meaning that the sample contained three latent groups of women distinguishable in terms of the severity and nature of their mental health needs.

**Table 26**

*Class Proportions (n/%) for All Models 1 through 4 with Two- & Three-Class Solutions*

Class Solutions & Classes: Temporary class labels	<i>n (%)</i>		
	Model 1 ( <i>N</i> = 920)	Model 2 & 3 ( <i>n</i> = 625) <sup>a</sup>	Model 4 ( <i>n</i> = 625)
<i>Two Class Solution</i>			
C1: Severe needs	275 (29.0)	191 (30.5)	191 (30.5)
C2: Limited needs	645 (70.1)	453 (69.5)	435 (69.5)
<i>Three Class Solution</i>			
C1: Severe needs	275 (29.9)	191 (30.5)	191 (30.5)
C2: Moderate needs/Severe self-harm	147 (15.0)	<b>109 (17.4)*</b>	<b>121 (19.3)*</b>
C3: Low needs	498 (54.1)	<b>326 (52.1)*</b>	<b>314 (50.2)*</b>

*Note.* This table displays two- and three-class solutions for models 1 through 4 and the number (%) of women assigned to each class. Proportions for the rejected and non-converging four-class solutions are not included here as they were found to be unreliable and misleading; a common occurrence in unstable, misspecified, and non-convergent models. Significant differences in proportions are presented in **bold font** followed by an asterix (\*) <sup>a</sup> Models 2 and 3 are identical in terms of class proportions (i.e., the addition of three auxiliaries in Model 3 does not impact class formation or model fit).

Second, the largest proportion of women in the overall sample (*N* = 920) were characterized by relatively low mental health needs (>50%). Third, a non-negligible proportion (≈ 30%) of women suffered severe mental health needs. Finally, most strikingly, a third, modest but non-negligible (14%) class of women had reduced overall mental health needs (e.g., few internalizing problems) but nevertheless maintained extremely high rates of self-harm and suicidal ideation and behaviours as compared to all other women in the sample.

**Model 2: The Addition of the Pre-Screen Risk Score Covariate**

Following careful inspection of all solutions tested with the least complex model (i.e., Model 1), the second stage of the LCA procedure was initiated; the Pre-Screen risk covariate was added to Model 2, and single-, two-, three-, and four-class solutions were again examined. Recall, due to missing data on the covariate, the sample was now reduced to 625 women. The addition of the covariate vastly improved model fit for the two- and three-class solutions, while

the single-class and the four-class structures remained unsupported (refer to summary Table 24 above).

Again, the three-class solution remained the best fit for the data. For example, both LRT were highly significant and favoured the  $k$  over the  $k-1$  model solution (i.e., the three- over the two-class structure). Classification accuracy (posterior probabilities) were favourable and none came close to falling below the  $\geq .70$  minimum recommended threshold for adequate posterior probabilities (range = .76 to 1.00). Further, the class proportions were affected by the covariate, such that they came to represent slightly more balanced subgroups than previously (refer to Table 26 above). The probabilities of the presence of each mental health indicator across classes did not change in any meaningful way between Models 1 and 2.

A four-class solution was tested as a matter of formality; this solution again resulted in model non-convergence because of model misspecification. Four latent classes simply were not supported by the observed data. In conclusion, accounting for the influence of the Pre-Screen risk score on the initial formation of the classes by including it as a covariate in the model substantially improved the overall fit of the model; in all other ways, the optimal solution (i.e., three latent classes) remained the same as that for Model 1.

### **Model 2 Summary**

Taken together, findings obtained in analyses of Model 2 can be summarized in four short take-home messages. First, as hypothesized, the addition of the Pre-Screen risk score covariate improved model fit substantially (refer to summary Table 25 above), and contributed modestly to class formation (i.e., class proportions changed modestly). Second, the optimal structure obtained in Model 1 – the three-class solution- also remained the favoured solution in Model 2. Third, the nature of the three latent classes remained virtually identical from Model 1 to

2; Class 1 consisted of severe mental health needs women, Class 2 consisted of women with moderate overall mental health needs but exceedingly high rates of self-harm and suicidal thoughts and behaviours (i.e., externalizing behaviours), and lastly, Class 3 remained the low mental health needs group.

### **Model 3: Using Auxiliaries to Better Understand Latent Classes**

Model 3 was tested next with one-, two-, three-, and four-class solutions. Importantly, there were no differences between Model 2 and Model 3, with the single exception of including three auxiliary variables; ACEs, age, and race. Importantly, auxiliary variables do not impact model fit, class solutions, class proportions, posterior probabilities (i.e., classification accuracy), or probability scales - they simply provide additional information pertaining to the already formally derived classes. Thus, a detailed discussions of model fit and model parameters would not be informative. Fit indices, entropy, and LRTs are presented together for Model 2 and Model 3 in summary Table 25 above. To reiterate, the three-class solution was again the most favorable latent structure, all aspects of model and class fit taken together, and the four-class solution again resulted in non-convergence and was rejected.

### **The Auxiliaries**

Next, the classes were examined in terms of any statistically significant differences on mean scores on ACEs, age, and race. Table 27 presents class comparisons on the three auxiliary variables. Of note, Mplus does not allow for specification of categorical auxiliary variables. Thus, Chi-Squares mean score difference tests meant for continuous variables were applied to the categorical variable race. This can under certain circumstances lead to unreliable test results. Thus, to guard against erroneous conclusions regarding any class differences on race, a binary race variable (*White vs. non-White*) was also computed using SPSS, and Chi-Square significance

tests suitable for 2x2 contingency tables were run to confirm initial auxiliary findings. Results are reported at the end of the below section outlining findings with auxiliary variables.

The overall Chi-Square was significant for this measure, meaning that ACEs significantly distinguished between classes (see Table 26 for Means, *SDs*, and results of pairwise and overall Chi-Square Difference Tests). Class 2 had the highest mean score ( $M = 3.92$ ,  $SE = 0.22$ ), followed closely by Class 1 ( $M = 3.78$ ,  $SE = 0.17$ ). The difference between Class 1 and Class 2 was not statistically significant. Class 3 had the lowest average ACEs score ( $M = 2.79$ ,  $SE = 0.14$ ). Pairwise Chi-square tests revealed significant class differences on ACEs for Classes 1 and 2 vs. Class 3. In conclusion, women belonging to the two classes with severe and moderate mental health needs had significantly higher average scores on childhood adversity than the low mental health needs Class 3 women.

There were no significant differences between classes on age or in terms of race, suggesting there is no relationship between class membership – or understood differently – the severity of mental health needs- and these demographics. As noted above, Mplus does not allow for specification of an auxiliary variable as categorical. Thus, race was included as a continuous auxiliary variable in these Chi-Square mean difference tests, potentially leading to inaccurate statistics and conclusions regarding any class difference on this variable. As mentioned, a binary race variable was created in SPSS and pairwise Chi-Square tests were conducted to compare classes on the binary race variable, measured as *White* vs. *Non-white*. The proportion of each of the three classes that were White ranged from 87.5% to 91.1%. Pairwise comparisons conducted in SPSS were converted into more easily interpreted Cohen's *d* effect sizes representing each class compared to one of the other two classes.

When Class 1 and Class 2 were compared, a Cohen's  $d$  value of 0.01 (95% CI; [-0.22, 0.23]) resulted, confirming the lack of any meaningful difference on race between Class 1 and 2. When Class 1 and Class 3 were compared, a Cohen's  $d$  of 0.11 (95% CI [-0.06, 0.28]) resulted, again supporting the initial conclusion that race did not differ across classes. Finally, comparing Class 2 and Class 3, a Cohen's  $d$  of 0.09 (95% CI [-0.09, 0.28]) resulted. Of note, Cohen's  $d$  values below 0.20 (Cohen, 1992) are by convention considered very small, and as is clear from the accompanying 95% confidence intervals, all three effects failed to reach statistical significance. In conclusion, class comparisons for race as a binary variable supported initial Mplus LCA findings for the original auxiliary variable; classes did not differ on race.

#### **Model 4: The Complex Final Model**

Finally, the fourth and final Model, subjected to the same potential class structures as the previous three models, was examined. In Model 4, the distal outcome of recidivism at 24-months follow-up was added. Despite low base rates, the shortest follow-up period was chosen as the outcome in Model 4 to preserve sample size. See Appendix L for a brief overview of alternative Model 4 results when the 36-month follow-up recidivism outcome was used in place of the original 24-month outcome: No meaningful differences between models resulted.

Predictably, the three-class solution was confirmed as the optimal solution, fit indices and LRTs pointing to improved fit over the two-factor structure (refer to summary Table 25 above). Notably though, the increased complexity of Model 4 was associated with a very modest drop in overall model fit (see summary Table 25 above) relative to the three-class solution for Models 2 and 3. The four-class solution was tested again and was once again rejected on account of model non-convergence.

**Table 27**

*Auxiliary Variables ACEs, Age, & Race in Models 3 and 4: M (SE) and Chi-Square Difference Tests (n = 625)*

Classes & Comparisons	Model 3: <i>M (SE)/ <math>\chi^2</math> value (p-level)</i>		
	ACES <sup>a</sup>	Age	Race
<i>Two-Class Solution</i>			
Class 1	3.78 (0.17)	34.09 (0.56)	1.28 (0.07)
Class 2	3.09 (0.12)	34.24 (0.41)	1.33 (0.05)
C1 vs. C2	<b>11.10 (p = .001)</b>	0.05 (p = .828)	0.35 (p = .552)
<i>Three-Class Solution</i>			
Class 1	3.78 (0.17)	34.09 (0.56)	34.09 (0.56)
Class 2	3.92 (0.22)	33.98 (0.80)	33.98 (0.80)
Class 3	2.75 (0.14)	34.35 (0.51)	34.34 (0.51)
C1 vs. C2	0.24 (p = .624)	0.01 (p = .912)	0.01 (p = .912)
C1 vs. C3	<b>21.56 (p &lt; .001)</b>	0.12 (p = .735)	0.12 (p = .735)
C2 vs. C3	<b>17.78 (p &lt; .001)</b>	0.13 (p = .717)	0.13 (p = .717)
Overall $\chi^2$	<b>30.62 (p &lt; .001)</b>	0.20 (p = .905)	0.20 (p = .905)
Classes & Comparisons	Model 4: <i>M (SE)/ <math>\chi^2</math> value (p-level)</i>		
	ACES <sup>a</sup>	Age	Race
<i>Two-Class Solution</i>			
Class 1	3.78 (0.17)	34.09 (0.56)	1.28 (0.07)
Class 2	3.09 (0.12)	34.24 (0.41)	1.33 (0.05)
C1 vs. C2	<b>11.10 (p = .001)</b>	0.05 (p = .828)	0.35 (p = .552)
<i>Three-Class Solution</i>			
Class 1	3.78 (0.17)	34.09 (0.56)	1.28 (0.07)
Class 2	2.73 (0.15)	34.42 (0.53)	1.29 (0.10)
Class 3	3.88 (0.22)	33.85 (0.77)	1.34 (0.06)
C1 vs. C2	0.12 (p = .731)	0.06 (p = .803)	0.006 (p = .939)
C1 vs. C3	<b>21.78 (p &lt; .001)</b>	0.19 (p = .665)	0.52 (p = .472)
C2 vs. C3	<b>17.11 (p &lt; .001)</b>	0.33 (p = .566)	0.21 (p = .645)
Overall $\chi^2$	<b>34.05 (p &lt; .001)</b>	0.47 (p = .790)	0.54 (p = .654)

*Note.* Rows denoting the two classes involved in a pairwise comparison present the obtained Chi-Square statistic and significance level.

Overall  $\chi^2$  = overall difference between all classes. <sup>a</sup> ACEs = Adverse Childhood Experiences Scale. Significant class differences are **bolded**.

With the inclusion of recidivism, the class proportions again changed to a modest extent, with Class 1 remaining static (i.e., identical to in Model 3 with a three-class solution), but with 12 women previously belonging to Class 3 now re-classified as Class 2 (refer to Table 26 above for class proportions). The probability of indicator endorsement within classes did not change in any meaningful way. Nevertheless, the probability scale for the complete, final Model 4 is presented in Table 28 below. Figure 6 below is a graphical representation of the mental health indicators probabilities across the three latent classes.

**Table 28**

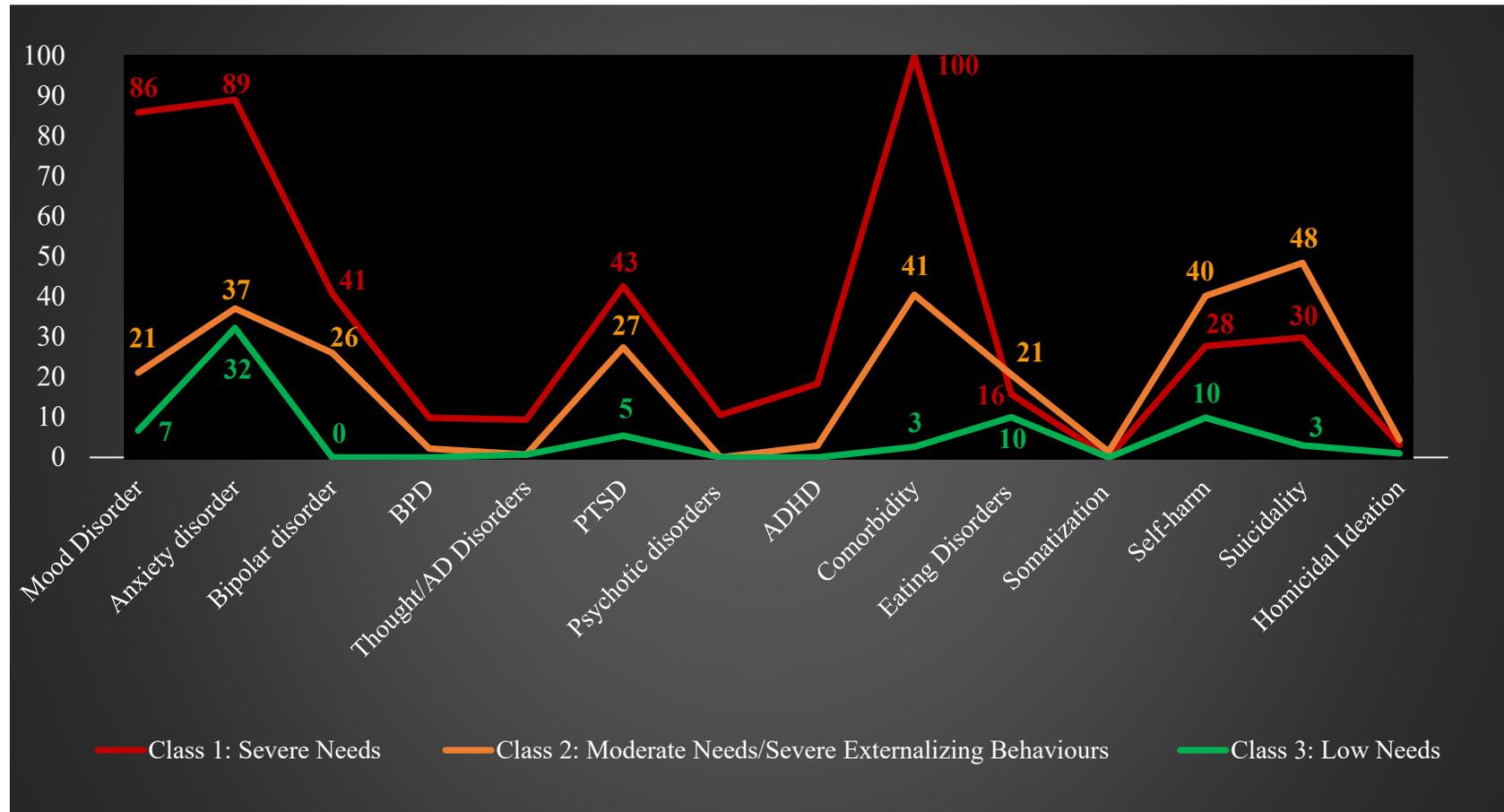
*Model 4 with Three Classes: Mental Health Indicator Probabilities (n = 625)*

Mental Health Indicators	<i>Mental Health Indicator Probabilities (%)</i>		
	Class 1 <i>Severe Needs</i> (n = 191)	Class 2 <i>Moderate Needs/ Severe Externalizing</i> (n = 121)	Class 3 <i>Low Needs</i> (n = 313)
<i>Current Mental Disorder Diagnoses</i>			
Mood Disorder	<b>85.9***</b>	21.1***	6.7***
Anxiety disorder	<b>89.0***</b>	37.1***	32.2***
Bipolar disorder	<b>40.8***</b>	26.0***	0.0
Borderline Personality disorder	<b>9.9***</b>	2.2↓	0.0
Thought & adjustment Disorders	<b>9.4***</b>	0.7	0.7
Post-traumatic stress disorder	<b>42.6***</b>	27.4***	5.4***
Psychotic disorders	<b>10.5***</b>	0.0	0.0
Attention-deficit/hyperactivity disorder	<b>18.3***</b>	2.9↓	0.0
Comorbidity	<b>100</b>	40.5	2.6
<i>Externalizing/Behavioural Indicators</i>			
Eating Disorders	15.7***	<b>20.6***</b>	10.0***
Somatization	0.0	<b>1.4</b>	0.0
Self-harm (Past/current)	27.7***	<b>40.2***</b>	9.9***
Suicidality (Ideation/attempts; Past/current)	29.8***	<b>48.4***</b>	3.0
Homicidal Ideation (Current)	3.1*	<b>4.3*</b>	1.0

*Note.* **Bolded percentages** indicate the class had the highest probable indicator endorsement rates. Significant class differences in endorsement rates: ↓ ( $p < .10$ ) \* ( $p < .05$ ) \*\*\* ( $p < .001$ )

**Figure 6**

*Model 4: Mental Health Indicator Probabilities (Percentages): The Final Three-Class Solution*



*Note.* This graph displays the probability scales in percentage across the 14 mental health indicators for each of the three classes. BPD = Borderline Personality disorder. AD = Adjustment; PTSD = Post-traumatic stress disorder; ADHD = Attention-deficit/hyperactivity disorder. Self-harm = past/current, Suicidality = suicidal ideation/attempts; past/current.

### **Post-Hoc Exploratory Analyses**

Reporting guidelines for LCA and related analyses suggest that, whenever possible, following the derivation of the best fitting class structure, post-hoc analyses should be conducted to provide further information about each class. Ultimately, women's correctional profiles should be as complete as possible. In the current study, two sets of supplementary analyses were conducted. First, pairwise comparisons between the three classes were conducted in order to determine whether the observed differences in indicator endorsement rates were statistically significant and whether they were of a magnitude likely to carry some clinical or practical meaning. Second, pairwise comparisons between the three classes were conducted with four distinct recidivism outcomes of varying follow-up periods; 24, 36, 48, and 60 months. In the sections below, the outcome of the pairwise comparisons on mental health indicators are presented first, followed by results of recidivism analyses.

#### **Pairwise Comparisons: Mental Health Indicators**

The initially reported differences in proportions of women with the given mental health indicators provided a general overview of the clear and major systematic differences between classes. However, the statistical significance of these and less apparent differences required further examination using pairwise comparisons of the latent classes. I conducted simple Chi-square significance tests based on cell frequencies (i.e., 2-by-2 contingency tables) for each indicator. In order to conduct these tests, it was necessary to create a new dataset in SPSS based on data imported directly from Mplus. Most importantly, the classification variable-the dichotomized classification variable that allows for the determination of each individual case's class membership- had to be imported. This variable was used to create filters narrowing data to be included in analyses to two of three classes for the purposes of pairwise comparisons of all

binary indicators. Data from the final complete Model 4 with the selected three-class solution was used to create the dataset. The Chi-square statistics resulting from initial analyses were converted into the most easily interpreted and frequently relied upon effect size measure in the contemporary psychology literature (Olivier et al., 2017) – the Standard Mean Difference (SMD)- most commonly referred to as Cohen's  $d$  (Cohen, 1992). Accurate conversions were ensured by using the very well-established Campbell Collaboration's online effect size calculator and converter tool (Wilson, 2001). This calculator allows for direct conversions of the original test-statistics into Cohen's  $d$  effect sizes along with their associated 95% confidence intervals. Despite the favorable official reviews of the calculator, the accuracy of the tool was verified by performing equivalent step-by-step computations using general formulae in Microsoft Excel. All tests conducted in this fashion confirmed the initial conversions obtained with the online conversion tool.

There are broad standards of interpretation of the magnitude of Cohen's  $d$  effect sizes commonly endorsed and frequently applied in contemporary research, including correctional psychology, available in the literature. Briefly, Cohen's  $d$  values of 0.20, 0.50, and 0.80, are considered small, moderate, and large effects, respectively (Cohen, 1992). A Cohen's  $d$  of 0.0 would of course indicate the total absence of any effect. Despite the appeal and obvious utility of these general rules of interpretation, they are meant to be employed alongside relevant information about the area of study and the conditions under which the effect size was derived. Base rate, for example, can affect the meaning of both Cohen's  $d$  and corresponding correlation coefficients ( $r$ ). For example, in certain areas of investigation, even very small effect sizes can have important practical or clinical implications (Fritz et al., 2012; Lee, 2016; Nakagawa & Cuthill, 2007; Vaacha-Haase & Thompson, 2004). The direction of effects (i.e., whether the

effect size is negative or positive) depends only on what group is chosen as the reference category. In a comparison of Classes 1 and 2, if Class 1 was the reference category and a significant positive effect size resulted, it would mean that Class 1 had a significantly higher proportion of women endorsing the indicator of interest than Class 2, provided the confidence intervals (CIs) did not include 0 (zero). Confidence intervals that include zero indicate that the difference as indexed by the effect size, is not statistically significant. The precise meaning of 95% confidence intervals is sometimes misunderstood. 95% CIs signify that if the range was calculated 100 times, 95 (i.e., 95%) of the resulting CIs would contain the true population effect size (Nakagawa & Cuthill, 2007).

Table 29 below presents effect sizes for all pairwise comparisons on mental health indicators and the co-occurring disorders variable computed post-hoc in order to further illuminate the complexity of mental health problems. The original suicidality indicator in the LCA models encompassed thoughts and attempts, effect sizes were also estimated for thoughts and attempts separately in these post-hoc analyses. The obtained effect sizes generally supported original probability scales for most indicators (Table 28). Nevertheless, some findings are worth noting here. First, eight of the 14 indicators differed significantly in all pairwise comparisons. Largest in magnitude were effects for the most commonly occurring internalizing conditions: mood, anxiety, and bipolar disorders. Class 1 had overwhelmingly high rates of these indicators relative to the other classes. PTSD also differentiated Class 1 from Class 3 with a very large, significant effect size ( $d = 1.17$ ) highlighting the high rates of trauma exposure in this Class. Class 2 also had high PTSD rates. Maintaining focus on large effects, Class 2 – the class characterized by the highest rates of externalizing behaviours, producing significant, moderate to

large, effect sizes, especially as compared to Class 3 ( $d > 1.0$ ) for self-harm and suicidality, with the largest Cohen's  $d$  of 1.56 representing differences between classes on suicidal behaviour.

Moreover, effect sizes comparing classes on eating disorders and somatization, as well as homicidal ideation were only significant in comparisons of Class 2 and Class 3. The existing comorbidity indicator and the co-occurring disorders variable added post-hoc aided in identifying Class 1 – “severe mental health needs” – both as the class with high rates on most indicators and the most complex mental health challenges. Co-occurrence significantly differentiated Class 1 from the remaining classes. In conclusion, the pairwise comparisons confirmed and further illuminated class differences in a way that promoted a comprehensive understanding of the challenges faced by the women belonging to each of these classes.

#### **Linking Mental Health and Recidivism: Additional Class Comparisons**

The final set of pairwise comparisons were conducted in order to examine potential class differences in recidivism at 24-, 36-, 48-, and 60-months follow-up. Table 30 below displays the number and proportions of women who recidivated in each class, and Cohen's  $d$  effect sizes and their 95% confidence intervals for each pairwise comparison. Chi-Square significance tests were utilized to examine the relationship between two classes and the binary recidivism variables, and the online calculator described above was used to convert the original coefficients into Cohen's  $d$  effect sizes.

Across all follow-up times, the highest recidivism rates occurred for Class 2-“Moderate needs and severe externalizing behaviour”, perhaps suggesting that the types of mental health challenges faced by women in this class were more relevant to recidivism outcomes than the mental health disorders most common in Class 1- the class with the most severe and complex mental health needs

**Table 29**

*Pairwise Comparisons of the Three Latent Classes on the 14 Mental Health Indicators & Variables of Interest Post-hoc (n = 625)*

Mental Health Indicators	n/(%)			Cohen's <i>d</i> (95% CI)		
	Class 1 <i>Severe Needs</i> (n = 191)	Class 2 <i>Moderate Needs/Severe Externalizing Needs</i> (n = 121)	Class 3 <i>Low Needs</i> (n = 313)	C1 <sup>a</sup> vs. C2 (n = 312)	C1 <sup>a</sup> vs. C3 (n = 504)	C2 <sup>a</sup> vs. C3 (n = 434)
Mood disorder	164 (85.9)	26 (21.5)	23 (7.3)	<b>1.68 (1.39, 1.97)</b>	<b>2.56 (2.28, 2.85)</b>	<b>0.41 (0.22, 0.60)</b>
Bipolar disorder	78 (40.8)	36 (29.8)	0 (0)	0.23 (0.00, 0.45)	<b>1.33 (1.21, 1.54)</b>	<b>1.10 (0.89, 1.32)</b>
Anxiety disorder	170 (89.0)	47 (38.8)	100 (31.9)	1.25 (0.99, 1.51)	<b>1.31 (1.10, 1.52)</b>	0.13 (-0.06, 0.32)
Thought/ad Disorders	18 (9.4)	1 (0.8)	2 (0.6)	<b>0.36 (0.13, 0.58)</b>	<b>0.45 (0.27, 0.63)</b>	0.02 (-0.17, 0.21)
BPD	19 (9.9)	3 (2.5)	0 (0)	<b>0.29 (0.06, 0.51)</b>	<b>0.52 (0.34, 0.70)</b>	<b>0.27 (0.08, 0.46)</b>
PTSD	91 (47.6)	38 (31.4)	16 (5.1)	<b>0.33 (0.10, 0.55)</b>	<b>1.17 (0.97, 1.37)</b>	<b>0.76 (0.56, 0.97)</b>
Psychotic disorders	20 (10.5)	0 (0.0)	0 (0.0)	<b>0.43 (0.20, 0.65)</b>	<b>0.54 (0.36, 0.72)</b>	-
ADHD	35 (18.3)	4 (3.3)	0 (0.0)	<b>0.45 (0.23, 0.68)</b>	<b>0.74 (0.56, 0.93)</b>	<b>0.31 (0.21, 0.50)</b>
<i>Any mental disorder</i>	191 (100)	93 (76.9)	133 (42.5)	<b>0.86 (0.62, 1.10)</b>	<b>1.43 (1.22, 1.65)</b>	<b>0.65 (0.45, 0.88)</b>
<i>Comorbidity</i>	191 (100)	49 (40.5)	8 (2.6)	<b>1.90 (1.59, 2.20)</b>	<b>7.60 (6.92, 8.29)</b>	<b>1.17 (0.95, 1.38)</b>
<i>Co-occurring disorders</i>						
<i>2-3 disorders</i>	128 (67.0)	49 (40.5)	8 (2.6)	<b>0.54 (0.31, 0.77)</b>	<b>1.99 (1.74, 2.23)</b>	<b>1.17 (0.95, 1.38)</b>
<i>4-6 disorders</i>	63 (32.9)	0 (0.0)	0 (0.0)	-	-	-
Somatization	0 (0.0)	2 (1.7)	0 (0.0)	-0.20 (-0.02, 0.43)	-	<b>0.22 (0.03, 0.41)</b>
Eating disorders	30 (15.7)	27 (22.3)	31 (9.9)	-0.17 (-0.06, 0.39)	0.17 (-0.002, 0.35)	<b>0.22 (0.03, 0.41)</b>
Self-harm (Past/current)	53 (27.7)	53 (43.8)	32 (10.2)	<b>-0.33 (0.11, 0.56)</b>	<b>0.47 (0.29, 0.65)</b>	<b>0.82 (0.62, 1.02)</b>
Suicidality (Past/current)	57 (29.8)	71 (58.7)	4 (1.3)	<b>-0.61 (0.38, 0.85)</b>	<b>0.94 (0.75, 1.13)</b>	<b>1.89 (1.63, 2.15)</b>
<i>Thoughts</i>	6 (3.1)	13 (10.7)	2 (0.6)	<b>-0.31 (0.09, 0.54)</b>	<b>0.20 (0.02, 0.37)</b>	<b>0.51 (0.32, 0.70)</b>
<i>Attempts</i>	51 (26.7)	58 (47.9)	2 (0.6)	<b>-0.44 (0.22, 0.67)</b>	<b>0.90 (0.71, 1.10)</b>	<b>1.56 (1.31, 1.80)</b>
Homicidal Ideation (Current)	6 (3.1)	6 (5.0)	3 (1.0)	-0.66 (-0.13, 0.31)	0.16 (-0.01, 0.34)	<b>0.25 (0.06, 0.44)</b>

*Note.* This table displays frequencies for 14 mental health indicators and post-hoc variables (in *italics*). Thought/AD = Thought and adjustment Disorders. BPD = Borderline Personality disorder; PTSD = Post-traumatic stress disorder; ADHD = Attention-deficit/hyperactivity disorder. Co-occurring disorders = frequency of two to three and four to six co-occurring disorders. **Bolded Cohen's *d*** effect sizes are statistically significant ( $p < .05$ ). “-” = frequency = 0 for one class within comparison; no effect size could be computed.

Class 1 and Class 2- the severe and the moderate needs groups, respectively, did not differ significantly in terms of recidivism irrespective of follow-up times. It is worth noting however, that recidivism rates were consistently higher in Class 2. Pairwise comparisons of Class 2 (“Moderate needs and severe externalizing behaviours”) and Class 3 (“low mental health needs”) revealed, as anticipated, significantly higher recidivism rates for Class 2, although effect were modest in magnitude ( $d = 0.30$  to  $0.34$ ). Lastly, comparing Class 1 – the most severely disordered class to Class 3 – the class with few mental health needs across follow-up periods of 38, 48, and 60 months, revealed significantly but modestly higher rates among women in Class 1 ( $d = 0.24$  to  $0.33$ ), providing some support for the contention that mental health status may be risk-relevant. Class 1 and Class 3 effects only fell short of statistical significance when the shortest follow-up period of 24-months was used.

**Table 30**

*Comparing Latent Classes on Returns to Custody across Four Follow-up Times (n = 625)*

Follow-up (months)	n/N (%)			Cohen’s d (95% CI)		
	Class 1 <i>Severe Needs</i> (n = 191)	Class 2 <i>Mod Needs</i> (n = 121)	Class 3 <i>Low Needs</i> (n = 313)	C1 <sup>a</sup> vs. C2 (n = 312)	C1 <sup>a</sup> vs. C3 (n = 504)	C2 <sup>a</sup> vs. C3 (n = 434)
24 months (n = 150)	20/150 (13.3)	20/93 (21.5)	21/235 (8.9)	-0.19 (-0.03, 0.42)	0.15 (-0.03, 0.32)	<b>0.31</b> <b>(0.11, 0.50)</b>
36 months (n = 126)	32/126 (25.4)	26/79 (32.2)	33/175 (18.9)	-0.13 (-0.09, 0.35)	<b>0.24</b> <b>(0.06, 0.41)</b>	<b>0.30</b> <b>(0.11, 0.49)</b>
48 months (n = 110)	35/110 (31.8)	30/69 (43.5)	37/137 (27.0)	-0.18 (-0.04, 0.40)	<b>0.28</b> <b>(0.11, 0.46)</b>	<b>0.34</b> <b>(0.15, 0.53)</b>
60 months (n = 98)	42/98 (42.9)	31/61 (50.8)	43/110 (39.1)	-0.11 (-0.11, 0.33)	<b>0.33</b> <b>(0.15, 0.50)</b>	<b>0.32</b> <b>(0.13, 0.51)</b>

*Note.* General recidivism included new offenses and technical violations resulting in return to custody to a Maine state prison. Mod = Moderate mental health needs and severe externalizing behaviours. **Bolded effect sizes** = significant ( $p < .05$ ). <sup>a</sup> This class was used as the reference category in relevant analyses; negative Cohen’s  $d$  values indicate that this class had lower recidivism rates than the comparison class.

### **Comprehensive Class Descriptions: Who Are the Women from Maine**

As described in detail in the preceding sections, across all four LCA models, three latent classes best described the women's mental health profiles. Together, the main LCA as well as key supplementary post-hoc analyses have provided relatively comprehensive characterizations of the three latent mental health classes. The following section will synthesize this information broadly, class by class, to facilitate an clear understanding of the differing mental health profiles of the women from Maine. The descriptions are based solely on the fourth and final Model (i.e., Model 4) with the three-class solution and relevant findings from supplementary post-hoc analyses.

#### **Class 1: Severe needs**

This class consisted of 191 women, representing 30.6% of the reduced LCA sample ( $n = 625$ ). The most prominent features of this class and those that distinguished it most clearly from both Class 2 and Class 3, were the very high rates of most internalizing disorders and the relative complexity of needs as demonstrated by a 100% comorbidity rate, as well as larger numbers of women having more than two current co-occurring disorders (33%). Among women in this class, mood, bipolar, and anxiety disorders, as well as PTSD (50%) and ADHD were very common. Relative to the remaining two classes, rates of relatively rare disorders such as thought and adjustment Disorders and psychotic disorders, as well as BPD, were also markedly elevated (i.e., roughly 10% vs. 0.0% to 3.3%). Higher rates of adverse childhood experiences as well as significantly higher recidivism rates at three of four follow-up times clearly differentiated Class 1 from Class 3- the low mental health needs women. However, when Class 1 and Class 2 were compared, these differences could not be considered meaningful and did not meet the threshold for statistical significance.

**Class 2: Moderate needs with severe externalizing behaviours**

Class 2 consisted of 121 women, (19.36%) of the overall sample of 625. While rates of internalizing behaviours were substantial, they paled in comparison to the severely disordered Class 1. Some degree of complexity of diagnoses did occur, with 40.5% suffering comorbid disorders. However, these internalizing disorders were not what set Class 2 apart from the other two classes. Rather, women in this class evidenced extremely high rates of all available indicators of externalizing behaviors, most relevant among them self-harm (44%), suicidal ideation, and past suicide attempts (47.9%), but also indicators such as diagnosed eating disorders (roughly 20%), and homicidal ideation. It is important to note here that the proportion of women reporting current homicidal ideation was still very low in this class. Nevertheless, the percentage exceeded both remaining classes (i.e., 5% vs 1% to 3.1%). Not surprisingly, in terms of adverse childhood experiences (i.e., potential childhood trauma), Class 2 was comparable to the severely disordered Class 1, but different significantly in the expected direction from Class 3—the low mental health needs class. Notably, in term of recidivism outcomes, women in this class had the highest recidivism rates across follow-up periods and classes; significantly and meaningfully higher rates emerged as compared to women in Class 3 (low mental health needs). Rates also exceeded those found for Class 1, though these differences fell short of statistical significance.

**Class 3: Low needs**

Three-hundred and thirteen women, representing 52% of the LCA subsample ( $n = 625$ ), made up the largest class of justice-involved women in this study. In terms internalizing problems, rates of diagnosed mental disorders were very modest ( $< 7\%$ ) for all but the anxiety disorders diagnostic category, which affected just above 30% of women in this class. For the

remaining diagnostic indicators, rates were very low, falling below 7% for all remaining disorders. Both the rate of mental health challenges, as well as their complexity set this class distinctly apart from Classes 1 and 2; only 2.6% of low needs women had current comorbid disorders. Relative to the women in Classes 1 and 2, their rates of externalizing behaviours were also low across all relevant indicators, only 10% reported a current eating disorder, 10% reported ever engaging in self-harm, 2% had previously attempted suicide, and 1% reported current homicidal ideation. Women in this class reported statistically significantly lower rates of adverse childhood events compared to the remaining two classes. Not surprisingly given the low overall mental health needs and the very low rates of externalizing behaviours, Class 3 evidenced the lowest recidivism rates of all classes, with rates significantly lower than Class 2 across all recidivism outcomes and significantly lower than Class 1 for three out of the four follow-up times. Despite the overall low mental health needs of this class relative to their counterparts in Classes 1 and 2, it is important to note that mental health challenges were far from absent even for this class; about 43% of these women nevertheless currently met diagnostic criteria for at least one mental disorder.

### **Conclusion**

Latent class analyses revealed that the current sample of women incarcerated in Maine state prisons could be reliably distinguished into three clearly distinct latent profiles according to mental health status in a consistent, statistically significant, and clinically meaningful way. Despite over 50% of the women having profiles characterized by low needs in the area of mental health, the remaining roughly 50% struggled with severe, varied, and complex mental health challenges at the time of assessment, supporting the mounting empirical evidence suggesting that mental health may be among the most challenging aspects of contemporary correctional practice.

### Discussion

The main aims of the current study was to a) identify the latent mental health classes (i.e., profiles) in a large sample of adult women incarcerated in Maine state, and b) to investigate whether the class differences in mental health status corresponded with meaningful between-class variations in recidivism rates.

Preliminary analyses involving an overall sample of 920 women revealed that roughly 66% of the women met the diagnostic criteria for at least one current mental disorder. About 20% reported problems with self-harm and suicidal ideation and attempts. The findings of the main LCA suggested that three latent classes best described the women in this sample; the three-class structure produced superior model fit and better class distinctions that appeared to reflect clinically meaningful differences between the distinct profiles.

Class 1 “severe needs”, represented approximately 30% of the sample. This class was characterized by severe and complex mental health challenges, and the highest rates of childhood adversity in the sample. The rates of mood, anxiety, bipolar, PTSD, and ADHD were very high in this class (e.g., mood > 85%). Although not necessarily higher than expected in a correctional population overall, relative to the other two classes, diagnoses of BPD, thought and adjustment disorders, and psychotic disorders were also much more common in this than any other class. The proportion of women with two to three, and four to six comorbid current disorders far exceeded those in Class 2 and 3. In line with the hypothesis that mental health problems would be positively associated with recidivism, Class 1 demonstrated significantly higher recidivism rates than Class 3, which consisted of women with low needs.

Class 2 made up 20% of the overall sample and had moderate overall needs but also severe externalizing behaviours. The rates of internalizing disorders in this class were extremely

modest as compared to the severe needs Class 1. Yet, rates of all indicators conceptualized as externalizing behaviours (i.e., behavioural indicators of poor mental health) were greater in this class than either of the remaining two classes and would be considered extreme in most all populations. Women in Class 2 had the highest rates of eating disorders, somatization, self-harm (past and current), suicidal ideation (past and current; 11%), suicide attempts (48% vs. 27% in Class 1), and homicidal ideation (5%). Rates of childhood adversity were mirrored those of Class 1. Class 2 consistently demonstrated the highest rates of recidivism across all four follow-up periods, although only comparisons with the low needs Class 3 reached statistical significance. This finding is in line with the meta-analytic findings of Study 2 above, which suggested that it is specific aspects or types of mental health challenges that may be relevant to correctional outcomes, rather than mental health status as a unitary construct. Specifically, it appears that externalizing indicators of poor mental health (e.g., suicidality, self-harm) may tend to be more strongly related to recidivism than internalizing indicators such as anxiety disorders. Certainly, previous research has found that self-injury and suicide attempts are related to aggression and violence among women (O'Donnell et al., 2015).

Finally, Class 3 “low needs” included 50% of women and thus made up the largest proportion of the overall sample. This class had very low rates of internalizing and externalizing problems across all indicators<sup>23</sup>, suggesting that on average, they suffered few severe mental health challenges. They had significantly lower rates of exposure to adverse childhood events (ACEs) relative to both Class 1 and Class 2. Finally, this class demonstrated the lowest recidivism rates of all classes; rates were significantly lower as compared to both remaining classes.

---

<sup>23</sup> An exception was anxiety disorders; approximately 1/3<sup>rd</sup> of women in Class 3 met diagnostic criteria for a disorder in this category.

The detailed investigations into the exact nature of the derived latent classes resulted in several important overall conclusions. First, the largest proportion of women in this custody sample did not appear to face overwhelming mental health challenges. However, despite 50% of the sample appearing relatively mentally healthy, some 40% still met diagnostic criteria for one mental disorder. This finding echoes contemporary correctional commentary identifying mental health as a considerable challenge to both men and women's corrections today (Lord, 2008).

### **Limitations**

The current study has several limitations that must be addressed. Focusing first on key limitations related to inadequate data and reporting standards, Orbis Partners Inc, the providers of both the archival dataset and all related support documents did not have access to information that clarifying the sample's treatment status. In short, reflecting the inadequate reporting standards noted for the studies included in the meta-analysis described in Chapter 2 above, it is not known whether all or any proportion of the women in this sample had successfully completed appropriate correctional treatment programs (i.e., interventions targeting the most relevant needs at the optimal intensity) prior to release into the community. Absent this information we cannot determine whether mental health constitutes a gender-responsive risk factor for recidivism, a specific responsivity factor, or both. If findings were obtained for an untreated sample, this would support aspects of mental health as risk factors for recidivism. However, if findings were obtained exclusively for women treated prior to release, it may be more appropriate to consider mental health a responsivity factors, at least until further research would demonstrate an additional, direct link. In the latter case, the observed relationship to recidivism might simply reflect the severely mentally ill women's inability to fully participate in, and benefit from, the criminogenic-needs focused interventions. In summary, the absence of

adequate data on these three items likely lead to a poorer understanding of what disorders, under what circumstances could be considered relevant to recidivism prediction.

Further, the dataset itself contained some minor but persistent problems that in some cases resulted in imprecise estimates (e.g., frequency counts) that necessitated the exclusion of a small number of potentially informative mental health-related variables. Specifically, variables indicating whether or not a woman had received psychosocial treatment or a psychopharmaceutical intervention for her current mental disorder, as well as whether she was medication compliant were inconsistently utilized and coded by the original assessors, resulting in overall data on those variables that lacked the precision as to allow for inclusion in the current analyses. The dataset was scoured for contextually relevant additional variables that could be used in combination with existing items to clarify the coding strategies utilized between cases and items, but none were available. In conclusion, due to these obstacles the relevance of mental health treatments received for current diagnoses could not be determined. In addition, while the original SPIn-W Pre-screen scores were available for all 920 women in the original sample, Orbis Partners' original variables representing these scores had been computed using an alternative method of domain-level and total score computation. Orbis Partners chose to count items coded -1 -originally indicating missing data- as 0 (zero), meaning they were interpreted the same as a valid risk score of 0 (zero). This strategy resulted in Pre-screen total scores for all 920 women in the original sample. Conversely, the current author of this study chose, with the aim of increasing accuracy of resulting computations, to exclude women who had missing data on any of the individual Pre-screen items (i.e., cases for whom items were coded -1 were excluded). This, as mentioned above, resulted in the original sample of 920 women being reduced to 625 in all but initial analyses.

Other sample-related limitations included the under-representation of both low and high-risk cases. More than 80% of the current sample consisted of women classified as moderate risk on the full SPIn-W assessment, with approximately 15% and 5% of the original 920 women being classified as low- and high-risk, respectively. Future research would certainly benefit from examining the replicability of the current mental health profiles in low- and high-risk samples.

The definition of recidivism was resulted in several related limitations. First, recidivism was limited to returns to custody specifically to a Maine state prison. This means that any re-incarcerations to local or remand jails, as well as any returns to prisons outside Maine were not detected. Likely at least partially as a consequence of this narrow definition, the recidivism base rates were abnormally low across all follow-up times. The base rate fell just short of the minimum recommended rate for the conventional 24-month follow-up period (i.e., 10%). Longer follow-up times (i.e., 36, 48, and 60 months) were associated with considerable attrition. When base rates are lower than 10%, the reliability and validity of the findings of recidivism analyses may be jeopardized (Lynn-Stewart et al., 2019; Tabachnick & Fidell, 2019). Ensuring adequate base rates and an appropriate length of follow-up is an especially salient issue in women's correctional research; Previous research has demonstrated that low base rates and short follow-up periods increase the risk of unreliable, inaccurate findings of any recidivism analysis (Lynn-Stewart et al., 2019). Nevertheless, despite the seriousness of this limitation, every effort was made to capitalize on all available data and examine the consistency of findings as closely as possible. First, while the main and final LCA model primarily in focus in the current report utilized the 24-month recidivism outcome (i.e., with a base rate just below the minimum recommended level), the model was also tested with the 36-month recidivism outcome, without any favourable difference in overall model fit or class solution resulting. Most importantly, in

post-hoc analyses linking the three classes with subsequent recidivism rates, pairwise comparison were conducted across all four available follow-up times; 24, 36, 48, and 60 months. Remarkably consistent findings resulted across outcomes both in terms of the magnitude and significance of effect sizes. Thus, despite the less than ideal base rates and reduced sample sizes, additional analyses meant to help clarify the reliability suggested that the current findings could nevertheless be considered reasonably accurate.

The assessment of mental health status was limited to a relatively modest number of indicators. A number of potentially relevant aspects of mental health, such as dissociation, unstable identity, emotional dysregulation, and symptoms clusters considered relevant in most correctional contexts but that do not in and of themselves constitute a diagnosis (e.g., volatile pattern of hostile, angry, and sometimes violent ruminations, chronic over-controlled anger, and unpredictable, explosive attacks of rage), were not addressed. Symptoms that can occur on account of multiple distinct diagnoses, as well as in the absence of diagnoses, such as hallucinations, persecutory delusions, and chronically intense suspiciousness of others and distrust across contexts were also excluded. Any self-reported mental health concerns not reflecting the existing items were not considered an area of focus in the assessment.

Conversely, the relevance of some indicators that were included may be brought into question. Some appeared to have limited utility in that they reflect mental health problems known to be rare across populations and that had extremely low prevalence rates in the current sample (e.g., somatization,  $k = 2$ ). Along similar lines, some indicators without any intuitively identifiable risk-relevance (e.g., eating disorders) were included. It should be noted that these indicators and disorders may be important for case management purposes, however. On a final, brief note, one indicator-current homicidal ideation- appeared unlikely to illicit accurate self-

reported rates in the context of formal risk/need assessments that the women know can directly impact their security and risk classifications as well as their likelihood of being considered a violence risk. In turn, correctional outcomes such as conditional release may be negatively affected by truthful reporting of this information. In conclusion, it is not entirely clear how informative this indicator is, including in the current study.

Another potential complication associated with measuring prevalence rates of psychiatric disorders is the high potential for unreliability of diagnoses across individual assessors; the unreliability of psychiatric diagnoses is still considered a major problem in contemporary psychiatry (Aboraya et al., 2006). Assessors differ in terms of simple subjective interpretations of individual clients even when other potentially influential factors are held constant. The influence of personal and systemic biases are documented in the literature (Baglivio et al., 2016), and individual psychiatrists' degree of experience and skill levels vary. Further, the assessments, diagnostic manuals, clinical interview protocols, and interview techniques, are highly inconsistent across assessors. In the current study, extremely high rates of comorbidity of current mental disorders were uncovered; some women reported up to six current comorbid disorders. Although these findings could plausibly accurately reflect the degree of complexity of mental health challenges in the sample, as proposed in the current LCA analyses, it is also possible that these large numbers of diagnoses actually reflect, to some undeterminable extent, the variety of interpretations of the same underlying conditions across individual assessors, contexts, and timing (e.g., a woman may be diagnosed with BPD following her first hour-long assessment with a new psychiatrist, while another psychiatrist assessing her over period of a year might conclude her presentation falls more in line with substance use, complex PTSD, and ADHD diagnoses). In short, what was perceived in the current LCA as measures of the complexity of mental health

challenges, could plausibly be partially accounted for by the unreliability of psychiatric diagnoses. Unfortunately, this issue is difficult to avoid, as low diagnostic reliability continues to plague the field of psychiatry in its entirety (Aboraya et al., 2006).

A notable limitation in terms of the comprehensiveness of latent classes descriptions was the fact that risk on SPIn-W assessment domains, particularly those reflecting major criminogenic needs such as antisocial attitudes, criminal history, and antisocial associates were not included in pairwise comparisons of the three latent profiles. Investigating the full correctional profiles of the women included in each class would have been ideal. However, because the SPIn-W Pre-screen triage tool was utilized as a covariate in all but the initial Model 1 and impacted initial class formations themselves. Thus, comparing the classes on the Pre-Screen measure would be tautological- employing circular reasoning.

Lastly, the power may have been somewhat reduced in the current LCA. Contemporary recommendations with regards to ensuring adequate power in LCA suggest that given an average Cohen's *d* effect size of 0.80, approximately 500 cases would be required for acceptable power. In the current study, all but Model 1 ( $N = 920$ ) had a sample size limited to 625, which may be considered modest to moderate. However, given that the three-class solution also clearly emerged as the optimal structure in Model 1, which utilized the full sample of 920 women, coupled with the completely consistent and unambiguous replication this class structure across all subsequent models, it appears reasonable to conclude that the current findings do reflect good reliability and thus that power could be considered adequate.

### **Future Research**

Replication of the current LCA findings (i.e., the three classes structure or the three classes with additional classes added in the case of multiple novel mental health indicators and

crimino-genic needs measures) is an important future goal. Replication studies should utilize the largest sample possible, compare women on key extraneous factors (e.g., auxiliary analyses), and mindfully select samples that represent the general population of women across risk classifications. If the SPIn-W or its shorter triage version (i.e., Pre-screen) are used in future LCA research, every effort should be made at outset to maximize the number of cases with valid Pre-screen and/or full assessment risk scores. When a large proportion of the sample must be excluded on account of either invalid or missing data, whenever possible, steps should be taken to compare women with and without missing data on all key variables in the dataset and clearly report in the study whether missingness was systematic or random.

Future studies should attempt to address major limitations to the existing literature, including those of the current study, as well as replicate current findings with different samples of justice-involved adult women (e.g., differing in terms of risk level, offense histories [violent vs. non-violent], and community vs. custody settings). Importantly, the relationship between the obtained latent classes and recidivism should be examined relying on higher base rates, longer follow-up periods, and with more inclusive recidivism outcomes. These adjustments could be expected to provide more accurate rates of reoffending than those employed here (Lynn-Stewart et al., 2019). Efforts to assess a much larger number of potentially risk-relevant mental health constructs should be made, and latent classes should be described in as much detail as available data allow, promoting the development of comprehensive correctional profiles, which in turn would more clearly support the advancement of our understanding of the characteristics and needs of this population. A deeper understanding and more comprehensive knowledge base would facilitate potentially meaningful improvements in all aspect of women's corrections. Conducting research involving psychiatric diagnoses in the most responsible fashion would

ideally involve prospective longitudinal study designs in which the highest degree of consistency across assessments could be achieved.

Future latent class and profile modeling studies should aim to include the three crucial mental health indicators missing in the current study – psychosocial mental health treatment, psychopharmaceutical interventions, and medication compliance and to re-examine the relationship between individual disorders and institutional and community correctional outcomes. Further, whenever possible, a broader range of indicators should be examined; given the unreliability and scientific limitations of existing psychiatric diagnoses, future research may benefit from assessing a number of specific symptom clusters that have been at least theoretically linked with criminal and antisocial behaviour.

Amongst the most immediate aims of future research should be the examination of mental health profiles in the context of, or alongside, established criminogenic needs. Conducting latent class or profile analyses, and to a certain degree exploratory and confirmatory factor analyses including both mental health and criminogenic factors would result in more comprehensively informative correctional profiles. Along the same lines as including additional mental health variables and criminogenic need factors, examining adversity and trauma in adulthood may also be a worthwhile endeavor, as we now know that a single or a small number of adverse experiences do not have the same impact on mental health as prolonged, repetitive, and varied trauma exposure consistent across the lifetime (i.e., see PTSD vs. Complex trauma; Brown et al., 2021).

Finally, a point of interest in the current study was the extremely elevated rates of externalizing self-destructive behaviours such as self-harm in Class 2 “moderate needs with severe externalizing behaviours”, which overall mental health status was interpreted as reflecting

moderate needs. Recall that relative to the “severe needs” Class 1, this class had radically reduced rates of all mental disorders and overall less complex mental health challenges. Recall that no significant differences in childhood adversity were detected. The question thus remains, what extraneous factors explain these unusually high rates of deeply dysfunctional self-destructive behavioural manifestations? Future research should further investigate the etiology of chronic self-harm and persistent thoughts of suicide and suicide attempts, as well as other externalizing indicators like eating disorders and somatization, as well as their relationships with correctional outcomes, including recidivism.

### **Implications for Practice**

There are several key implications following from the current findings. Although mental health constitutes an enormous obstacle in correctional practice currently, allocating scarce mental health resources indiscriminately across the institutional population would likely lead to problematically low cost effectiveness, given that as much as 50% of women do not have overwhelming challenges in this area. There would likely be no identifiable benefits to either individual clients nor to correctional agencies of targeting women without significant mental health obstacles for mental health-focused interventions. Indeed, this would violate the key Risk principle (Andrew & Bonta, 2010 provide an overview of the basic principles of the RNR correctional model), which dictates that the highest risk clients should be assigned to the most intense interventions, while very low risk clients may be better served by low intensity or no treatment at all (Andrews & Bonta, 2010). Thus, early identification of clients with major mental health concerns is crucial, whether mental health constitutes a risk factor, a responsibility issue, or both.

Further, establishing a highly generalized approach to mental health may also result in sub-optimal outcomes, given that distinct profiles are associated with very different areas and degrees of dysfunction. A flexible, multi-modal model that can accommodate the unique needs of several distinct classes of women struggling with mental health would arguably have better odds of successful outcomes for the largest number of women. Chapter 4 below provides an elaboration on specific related recommendations.

Mental health needs and trauma histories characterize the lives of such a large proportion ( $\geq 50\%$ ) of this population that new treatment modalities alone may not suffice to meet their needs. Rather, mental health should to be considered at every stage and level of the correctional operation. First, as is already clear from the existing empirical literature, correctional management, front-line staff, and clients alike would benefit substantially by the consistent employment of holistic, trauma-informed approaches to individual client and overall population management (Brown et al., 2021; Covington & Bloom, 2008). Developing effective individual case management plans require an in-depth appreciation for the potential negative effects of severe mental disorder on client-case manager communication (i.e., rapport building), institutional behaviour, order, and safety, and treatment readiness. Case managers might develop more effective plans if a flexible, adaptive management scheme allowing individual woman's specific concerns to become part of the process of identification of major focus areas was embraced. Over-reliance on general approaches that ignore findings such as those reported in the current study; clear and unambiguous identification of distinct mental health profiles associated with unique sets of major mental health challenges, could result in ineffective methods resulting in unnecessary and ineffective waste of resources.

Adjustments to or the development of new correctional treatments modalities should be developed to address women's mental health needs. However, it has not yet been determined whether mental health is best understood as risk factor or as specific responsiveness, the most effective adjustments remain unclear until future research has adequately addressed this key outstanding issue. Chapter 4 below provides an in-depth discussion of viable practice options for either eventuality.

Policy and established practices currently guiding day-to-day institutional operations could be improved considerably by promoting knowledge and acceptance of the nature of the lives and the sometimes enormous challenges faced by individual women sentenced to long periods of incarceration. An adequate understanding and demonstrated respect could open the door to meaningfully improved responses to institutional challenges that more clearly fall in line with correctional agencies' key mandates, including safe and humane housing and treatment of clients in their custody and effective treatment interventions. Moreover, reductions in violent incidents between and within client and staff groups would improve institutional safety for all parties and facilitate improved correctional outcomes in general. Committing to developing alternative, evidence-based options to handle undesirable and disruptive institutional behaviours that are mental health-related should be a priority. Human rights advocates, legal scholars, and external oversight agencies such as the OCI (e.g., OCI, 2019) have long identified the need to relinquish the use of traditional responses to disruptions by mentally ill women (e.g., the routine use of long-term solitary confinement). As the empirical literature to date has demonstrated, such alternative measures help reduce the rates of re-traumatization and chronic, cyclic exacerbation of existing problems among already severely troubled clients (OCI, 2019).

## Chapter 4

### General Discussion

The current two-program of study addressed the debated issue of the potential relevance of gender in the etiology of criminal behaviour and the factors that affect the risk of recidivism among justice-involved men and women. It is important to realize that women involved in crime did not become a population of major theoretical or empirical research interest until relatively recently. Previous correctional research focused almost exclusively on men and results were assumed directly applicable to all correctional subpopulations, including adult women. Scholars aligned with traditional correctional perspectives such as the Risk-Need-Responsivity (RNR) model that dominates contemporary correctional research and practice have retained this contention, stating that women and men's motivations and pathways into criminal offending, as well as the specific environmental, psychosocial, and personality factors that contribute to those behaviours, are the same. In short, within this framework, gender is irrelevant to risk and needs assessments, and should not influence treatment approaches.

In contrast to these traditional views, gender-responsive (i.e., women-focused) scholars suggest that gender can be a highly relevant factor that must be considered in correctional research. Without considering gender, obtaining an adequate scientific understanding of the multi-factorial etiology of criminal behaviour among both genders, salient psychosocial risk factors, and the most appropriate treatment targets (i.e., needs), will remain beyond our reach. In the context of correctional research, an understanding of the role of gender is required for the development more humane and more effective correctional practice.

Gender-responsive scholars face multiple criticisms, including a lack of adequate theoretical models, an over-reliance on highly subjective and misleading methods and analyses

that fail to meet basic scientific standards (e.g., early qualitative pathways work; Chesney-Lind, 1989) and continued ignorance of what gender-neutral scholars consider overwhelming evidence supporting a very generalized, traditional, gender-neutral approach to corrections. Conversely, gender-responsive scholars object to the traditional scholars' willingness to apply their findings indiscriminately across populations, despite their investigations typically excluding women. Despite vast growth and expansion within contemporary gender-responsive scholarship in recent years, the available research is still very limited relative to the existing literature focused on men. Key to current gender-responsive scholarship is the investigation of potential gender-responsive risk factors; such factors can be either gender-specific (i.e., only relevant to women and not to men) or gender-salient (i.e., more strongly predictive of recidivism among women than among men). Commonly cited gender-responsive risk factors include romantic relationship dysfunction, complex trauma, low self-efficacy and self-worth, and mental health—the factor under investigation in the current program of research.

Research investigating the potential relevance of mental health to recidivism among women and the underlying mechanisms responsible for this potential relationship, is still in its infancy. The two studies constitute complementary elements of the current program of study aimed at addressing this specific gap in the empirical correctional research literature. The current research focused exclusively on the under-studied population of justice-involved adult women and more specifically on mental health in this population.

### **Summary of Main Findings**

Study 1 aimed to quantitatively synthesize all available data speaking to the potential relationship between mental health and recidivism in comprehensive meta-analysis; evaluating the evidence of the relative risk-relevance of mental health.

Three main take-home messages resulted from the comprehensive meta-analysis. First, the comprehensive literature review identified a dearth of empirical research on the relationship between mental health and recidivism among justice-involved adult women. First, the number of available independent effect sizes per mental health predictor was extremely modest. Second, a host of major limitations to methodology, analyses, and reporting standards characterized almost all included studies. The suboptimal number of available independent effects negatively affects statistical power; the current findings are in need of replication and future research endeavours should attempt to examine a broader range of mental health factors when possible. Researchers' willingness to take honest stock of our adherence to basic scientific principles to date, their willingness to attempt to address limitations of existing studies and embrace basic reporting guidelines will in large part be what determines the utility of future research reports.

Second, main findings indicated that the risk-relevance of mental health depends entirely on the exact problem examined; not all aspects of mental health are relevant to women's correctional outcomes. In the current study, meeting diagnostic criteria for at least one (i.e, any) mental disorder, depression, PTSD, and an extensive psychiatric history were significantly and modestly associated with recidivism. Conversely, anxiety, self-harm and suicidality, psychosis, and unspecified, mixed personality disorders were found to be unrelated to recidivism.

These findings do raise questions about gender-neutral scholar's claims that only risk factors supported in studies using all men samples can be relevant to women's recidivism. Indeed, although they must be considered tentatively until future replication can be achieved, current results consistently supported gender-responsive claims that several mental health factors are relevant to correctional outcomes among adult women.

Third, it is crucial to understand that the current meta-analytic findings cannot speak to the question of whether mental health is a direct risk factor for women or whether it constitutes a specific responsivity factor. To address this contentious issue, included studies' overall sample treatment status would have to be reported. Not a single author reported whether the women in their sample(s) had completed correctional treatment programs aimed at reducing recidivism prior to release into the community<sup>24</sup>. As a result, the connection between PTSD and recidivism, for example, may result from direct effects of PTSD symptoms on women's odds of recidivating (i.e., a risk factor) or it could reflect a reduction in treatment efficacy associated with PTSD symptoms (i.e., specific responsivity) such that their presence inhibits women with this diagnosis from fully participating in and benefitting from treatment.

Study 2 used LCA to identify the mental health profiles present in a large sample of adult women incarcerated in Maine. Latent classes or profiles are unobserved subgroups of women who are similar in terms of the nature, severity, and complexity of their challenges. Being able to identify such groups aids in determining the nature and severity of the mental health challenge posed by the population of adult justice-involved women, and ultimately in determining how to manage, support, and treat women in accordance with their specific challenges in correctional settings. Further, the ability to distinguish between women with and without special mental health needs also helps reduce unnecessary resource expenditures and unhelpful case management plans not necessary or helpful to mentally healthy women. In the LCA, indicators of mental health were extracted from the Service Planning Instrument for Women (SPIn-W;

---

<sup>24</sup> Of note, the archival dataset utilized in Study 2 also did not provide any information regarding treatment, preventing even informed speculation regarding this key issue.

Orbis Partners, 2006), a risk/need assessment and case management tool frequently utilized with adult women across North-American jurisdictions.

Study 2's primary findings of interest are provided briefly again here. First, the preliminary, LCA, and post-hoc analyses all revealed that there is a great degree of variability in mental health status of justice-involved women serving custody sentences; among the three reliably identified latent profiles uncovered in analyses, the largest proportion had very low overall mental health needs (50% of entire sample). Another 30% had extremely high rates of almost all potential mental health problems assessed. Further, comorbidity characterized every woman in this class with individuals meeting diagnostic criteria for between two to six co-occurring currently symptomatic mental disorders. In summary, the characteristics of women in with this profile suggested the presence of severe and complex mental health needs. The last class encompassed the final 20% of the sample. The profile of women in this class can be characterized as unique and atypical; their overall mental health needs fell in the moderate range. Yet, they had extremely high rates of severe behavioural dysfunctions including non-lethal self-harm and suicidality (i.e., a construct encompassing suicidal ideation and past suicide attempts) that far exceeded even women in Class 1 who had the most severe overall mental health needs. Approximately 44% to 59% engaged in self-harm or suffered suicidal ideation or past attempts. Finally, results of analyses comparing recidivism rates across classes consistently supported the relevance of mental health. Specifically, women belonging to Class 1 ("severe needs") as well as women in Class 2 ("moderate needs with severe externalizing behaviours") had similar recidivism rates, while the "low needs" class had significantly lower rates across four follow-up times compared to both classes with mental health concerns.

Study 2 constitutes an important contribution to existing literature in several ways. First, relative to previous LCA including mental health, which typically relied on one or a very small number of relevant indicators of mental health, the current study includes a wide variety of both internalizing disorders and externalizing behaviours, allowing for the construction of more comprehensive mental health profiles. The current study utilized a larger than average sample size and unlike previous research focused specifically on incarcerated adult women. The methodological rigor under which the latent mental health profiles were derived, and the thorough post-hoc investigations examining their meaningfulness (i.e., magnitudes and significance of effect sizes representing class differences on LCA indicators and other mental health variables constructed post-hoc) of apparent class differences constitutes an additional advantage. Lastly, the current study examined the link between class membership and multiple recidivism outcomes, further strengthening the study's contribution to the existing knowledge base and the practical utility of its findings. The study is especially relevant to the current correctional climate, in which the extremely high rates of mental disorder among adult women and the proportion considered significantly impaired is coming into focus. Regardless of gender, mental health is considered among the major challenging to effective correctional practice today (Lord, 2008).

Taken together, meta-analytic findings in Study 1 and results of LCA in Study 2 consistently support the contention that some aspects of mental health are relevant to the prediction of recidivism among adult women. As highlighted above, these mutually supportive findings do not specify whether the association is primarily direct or indirect. Although results may reflect mental health as a direct risk factor for recidivism in this population, the possibility remains that the association is due to the undermining effects of severe mental health related

dysfunction on women's ability to benefit from existing correctional interventions. Despite this unanswered question, the current results lend clear support for the gender-responsive contention that mental health is an important factor to consider in women's corrections, including in the context of treatment and assessment of recidivism risk. At the same time, the results are also fully compatible with traditional correctional perspectives conceptualizing mental health as a specific responsiveness concern.

### **Limitations of the Current Program of Study**

A notable limitation characterizing this area of empirical investigation generally and the current program of study specifically was the lack of adequate information available in the existing literature. The problem is multifaceted. First, the dearth of data is partially attributable to the very limited number of independent empirical studies dedicated to examining mental health in this specific correctional population. Second, the studies that do attempt to address this issue typically suffer limitations that unduly affect the reliability, validity, and utility of their findings. First, study designs are frequently severely flawed. Second, less than ideal sampling is frequently relied upon such that generalizability is severely reduced (e.g., studies attempt to speak to the issue as it pertains to entire adult woman population, yet rely on extremely unusual samples to do so). Third, a pervasive lack of appreciation for the potential influence of multiple contextual factors and a resulting failure to attend to them permeates the literature. Fourth, methodological concerns- including both within- and between-study definitional and measurement flaws and inconsistencies that impact the individual effect sizes such that they cannot be considered to reflect the same association or population across studies. Ultimately, this problematizes their aggregation in meta-analyses. Fifth, suboptimal analytical approaches that do not permit for the conclusions so commonly proposed based on their outcomes occur with some regularity. Finally,

potentially among the most damning shortcomings of much of the existing literature, is the persistent disregard for basic reporting guidelines, particularly in terms of what constitutes unacceptable omissions of crucial information. Consistent failure to report the data necessary to make the findings interpretable, meaningful, and potentially suitable to inform policy and practice reveals a far too permissive research culture. Recall the failure to mention treatment status in all 18 studies included in the meta-analysis. This prevented even informed speculation on the nature of the relationship between the mental health predictor and the recidivism outcome, let alone any identification of correctional approaches likely to improve the personal lives and community reintegration potentials of women suffering major mental health challenges.

The area of study in general further suffers from a lack of theoretical models that could account for the empirical association between certain mental disorders and recidivism in any population and in the justice-involved adult women population specifically. The current program of study could not resolve the atheoretical nature in existing literature or absolutely prevent any impact thereof on the current investigations, despite the significant effort made to utilize disorder-specific symptomology and related behavioural expressions to provide a minimal theoretical backdrop that could be utilized in attempts to explain the observed relationships between mental health and recidivism. It must be noted that part of the lack of adequate theoretical models is a natural consequence of the lack of knowledge in the field of psychiatry regarding the true etiology of the majority of their accepted diagnoses.

Despite every effort to identify and utilize any and all available indicators of mental health in both Study 1 and Study 2, both investigations fell short of this goal: A large number of potentially relevant aspects of mental health remains unaddressed. Study 1 could only address indicators for which an absolute minimum of four independent, univariate effect sizes with

general recidivism as the outcome were available in existing studies. Mental health indicators available in Study 2 were restricted to those included in the SPIn-W risk assessment measure (Orbis Partners, 2006), which, while remaining among the risk/need assessments most inclusive of gender-responsive risk factors, still leaves many aspects, indicators, and diagnoses unprobed. The mental health profiles uncovered in this study, though informative, consistent and reliable, and relatively inclusive compared to previous mental health LCA classes and profiles with both adult women and delinquent girls, cannot be considered to reflect a complete picture of justice-involved adult women's mental health challenges relevant to the correctional context.

Lastly, it is worth mentioning that statistical power to detect the true effect sizes with optimal precision was sometimes reduced by a non-negligible margin. In the meta-analyses, the reliability of the effect size, and the precision with which it can be estimated, naturally depends on the number of independent effects included. With the exception of the most commonly cited disorders of interest, only very few effects were available for each indicator. The results pertaining to these indicators should be regarded tentatively, clearly acknowledging their novel, exploratory nature without losing sight of the study's utility as a first step towards building a foundation for future scientific knowledge accumulation. In terms of Study 2, recommendations have been made to ensure adequate power in most analyses, and the application of these to the current investigation resulted in the conclusion that in some instances, power may have fell significantly short of ideal levels. However, some study features and results strengthened the reliability and validity of conclusions even in the face of power limitations, including complete consistency of the optimal class structures across all models, particularly in the initial model which sample size far exceeded established minimum standards. Thus, while this limitation

requires mention, it appears implausible to have affected analyses in ways that could invalidate the study's main findings.

### **Implications for Future Research**

Chapters 2 and 3 Discussion sections identify the implications of the current research findings and the limitations associated with them in detail. Nevertheless, a brief overview of the main implications for future research is presented here. First, and most obviously, extensive research efforts are still required to develop a deeper understanding of the relationship between mental health and recidivism among justice-involved adult women. There is a need for sound theoretical models and empirically substantiated insights that address the exact nature, intensity, and relevance of various mental health concerns to correctional outcomes across samples and contexts. Only with these building blocks in place can the development of fully informed, carefully constructed, effective approaches that can successfully improve the personal circumstances for the women facing overwhelming mental health challenges and conflicts with the law simultaneously as well as address mental health as a risk factor and/or responsivity issue with the aim of reducing recidivism. Quantity however, only furthers empirical knowledge in so far as researchers conducting the new research studies are willing to address all limitations under their control to the maximum extent possible. The utility of these future research endeavours depends however, on researchers dedication to developing sound research project and study designs, select samples with a view towards generalizability, the establishment of adequate and clearly stated definitions that can be employed across independent studies, address mental health more holistically, conduct group comparisons and other analyses which results apply directly to the relevant research questions, and engage in adequate reporting practices. Transparency should be the foremost guiding principle grounding the research endeavor as a whole, with reporting of

limitations being especially important. Every effort should be made to systematically collect, analyze, and report all relevant contextual information that could plausibly affect the outcome of analyses and the inferences that follow. In addition, analytic strategies and reporting practices should attempt to make findings accessible and comparable across studies (e.g., effect size measures should be selected with care and should allow for direct comparison with, or minimally for conversion into, the most commonly reported effect sizes to date).

Future studies should widen the assessment of mental health by including previously neglected symptom clusters, mental health related behavioural abnormalities, specific disorders previously neglected or insufficiently investigated, and other individual symptoms and facets of mental health believed to be potentially of special importance in correctional contexts, such as emotional dysregulation. In addition, potential indirect indicators of mental health status such as psychiatric history, involuntary psychiatric hospitalizations, community mental health supports, psychosocial treatments, and medication adherence as it related to specific mental disorders should be included in future research. Relatedly, studies developing mental health profiles for correctional use should attempt to include established risk/need factors (i.e., criminogenic needs), other gender-responsive risk factors (e.g., self-efficacy), and overall estimated risk of recidivism, allowing for the derivation of holistic, comprehensive correctional profiles. When recidivism is the outcome of interest, it should be clearly defined and selected with relevance to the correctional context of the specific setting and sample in mind. For example, with community samples, technical violations may be a major outcome of interest that could be examined separately from new offenses, base rates and sample sizes permitting. Conversely, for custody samples, re-incarcerations may be the main focus. Important to all recidivism studies, but of particular relevance in specific subpopulations with consistently low base rates such as adult

women is the use of follow-up periods long enough to allow for a resulting minimum recidivism base rate that allow for reliable outcomes of analyses (Lynn-Stewart et al., 2019). Regardless of the specific outcome and follow-up chosen, complete and clear reporting of the nature of the outcome should not be omitted.

Reports pertaining to investigations of mental health as a potential risk factor among adult women rarely if ever take steps to ensure group equivalency on relevant extraneous and contextual factors that could influence the likelihood of obtaining significant group differences and the magnitudes of those differences. In none of the 18 studies included in the Study 1 meta-analysis in which group comparisons were conducted were a-priori or post-hoc group equivalency addressed. At minimum, researchers should report the attempts made at ensuring equivalency either a-priori or post-hoc, clearly report what factors groups were compared on to obtain equivalency, identify any outstanding concerns or limitations associated with potentially systematic group differences with the power to obscure the relationships of interest, and fashion reports in such a way as to ensure transparency and consumer's ability to independently assess the soundness of the study.

As discussed in detail in Chapter 2, no studies identified for the meta-analysis, acknowledged the fundamental flaw common to the majority of investigations. Rather than a disordered group consisting of women with only one disorder and a comparison group of women with no diagnosed disorders, the comparison group typically consisted of all the women without that one specific the disorder of interest, including many with other disorders. Considering that rates of mental disorder are very high in the adult women population, comparisons are unlikely to reflect all healthy controls without special efforts made to ensure it. The assortment of other disorders present amongst controls and their potentially unique relationships to recidivism could

obscure results in all directions, from minimizing substantial effects and artificially inflating modest effects to the complete nullification of important effects in either direction. Because disorder prevalence rates are high, some effort may be required to secure an adequate subsample of women without a disorder. Using the current Study 2 as an example however, despite incredibly high rates of mental disorders in the sample, as many as 216 (35%) women had no diagnoses. Provided the minimum recidivism base rate is obtained, constructing a defensible comparison group for this study would not be an unreasonably difficult task. At minimum, researchers should report the nature of groups honestly and acknowledge potential for reduced reliability and validity when present.

### **Treatment efficacy: An Investigation to Consider**

Future research should investigate whether mental health is likely to be exclusively a responsivity issue. Informative would be a multi-pronged treatment outcome study comparing pre-treatment, post-treatment, and treatment change scores on accepted measures of main treatment targets multiple times over the course of treatment separately for women with and without the diagnosis or indicator under investigation. Ideally, a basic examination of group equivalency, and if necessary, efforts to adjust subsamples to achieve comparability on factors other than the disorder of interest would be made. Ideally, such studies would also assess clinically significant change – treatment change that demonstrates a change from scores or indicators falling in the dysfunctional range to scores falling in the functional range and vice versa (see Pettersen et al., 2015 for an example). The women would be followed for an adequate period- three to five years ideally- while at risk in the community before recidivism outcomes would be compared. Conducted with care, this approach would address differences in the magnitude of treatment change and the clinical significance of this change, between disordered

and non-disordered groups that may give some indication of whether the disorder itself affects women's ability to benefit fully from treatment. If disordered and non-disordered women evidenced equivalent treatment change, and both the same number of women with functional and dysfunctional scores were released from both groups, but the diagnosis under investigation still predicted recidivism, this would be an indication of the disorder being a risk factor alongside established criminogenic needs. If on the other hand, the disordered group demonstrated systematically reduced change relative to the healthy control group, and change scores predicted recidivism, this may be an indication that the disorder is part of a mental health responsiveness issue. Lastly, such treatment analyses should be complemented by comparisons of untreated disordered and non-disordered women. Although such research efforts take considerable time and requires a somewhat higher degree of research and statistical expertise than some simpler designs, they could lead to considerable advancements in our understanding of the role of mental health in women's corrections. The development of adequate institutional protocols, management, and treatment strategies for mentally ill women depends on this understanding.

### **Implications for Policy and Practice**

The main findings of the current program of study are undeniably relevant to correctional practice and could, provided current findings are replicated and extended as appropriate, be employed to aid in the purposeful, humane, and effective management of this correctional subpopulation that is characterized by extraordinarily high rates of severe and persistent mental health challenges. Mental health is currently recognized as a major obstacle to efficient, safe, and sustainable institutional management. Consistently reported complications of housing and managing a large number of severely mentally ill and therefore volatile women in an environment characterized by high stress, daily regimes and a culture focused on power and

control, frequent re-traumatizing situations, and psychological and physical violence include loss of institutional control by legitimate means, high degrees of stress and trauma among both staff and clients, frequent episodes of institutional misconduct, new offenses, and self-directed as well as interpersonal aggression<sup>25</sup> and violence. The loss of control by correctional authorities and the many high-risk situations that occur as a result tend to be responded to with punitive measures that further deepen hostilities and distrust between custodians and their clients. For example, acutely mentally ill women who engage in self-harm or suicide attempts are frequently transferred to administrative segregation cells characterized by social, material (including necessities such as hygiene items, clothing, and mattresses, blanket, or pillows for sleeping), and psychological deprivation and further trauma. Predictably, these conditions further exacerbate existing mental health problems leading to the problem behaviours in the first case. The riskiest most damaging scenarios are applications of this regime to women chronically engaged in severe self-harm and suicide attempts. The case of Ashley Smith, a 19-year old woman who died by suicide after spending four years in a segregation cell as a form of punishment for her self-harming and other disruptive behaviours (see Sapers, 2008 for a full report on this case). This practice is now considered an unambiguous, clear violation of human rights and indeed a form of formally recognized torture (Canadian Human Rights Committee [CHRC], 2012; United Nations General Assembly, 2011).

A host of external oversight agencies (e.g., the OCI), human rights groups (CHRC, 2012), the Canadian Bar Association (CBA), and international agencies such as the United Nations (UN) have consistently attempted - over the past 50 years - to exert pressure on

---

<sup>25</sup> In a study by Völlm & Dolan (2009), 45% of the women in their sample had histories of self-injury, and half of these started self-harming for the first time while incarcerated.

correctional agencies to adopt humane and evidence-based methods to safely manage mentally disordered clients, and this pressure has mounted in recent years. Women-focused correctional researchers have further identified a need for staff awareness of the lived experiences of justice-involved women, education on trauma and its effects on mental health, as well as on severe mental disorders and acute psychiatric crises in general. Concretely, the consistent recommendation is the adoption of trauma-informed care across correctional contexts, adequate knowledge of and basic training on how to utilize legal, humane, and evidence-based responses to mental health crises, and at least for a subset of women, the addition of emotion-focused therapy modules to existing cognitive-behavioural correctional treatment (Blanchette & Brown, 2006; Covington & Bloom, 2008). Current findings certainly do not contradict these recommendations.

Beyond institutional control, the findings of the current program of study consistently support a meaningful link between some mental health conditions and recidivism. Such findings, particularly if reliably replicated, may aid correctional authorities in terms of gaining a full appreciation for and understanding of the potential benefits of developing and adopting largely non-punitive, trauma and mental health–informed methods for coping with severe mentally ill women and women in psychiatric crises both in and out of the institutions. Provided with sufficient empirical evidence linking mental health to correctional outcomes like recidivism and institutional infractions, correctional authorities may come to view an investment in such adaptive responses as likely to benefit agencies, institutions, staff, and clients alike.

Effective adaptations must be evidence-based and contemporary interdisciplinary experts' recommendations for general practice should be carefully reviewed and considered. Naturally, there may be some questions regarding the exact nature of the most effective

adaptations and how to employ them given that little is yet known about the exact etiology and mechanism underlying the observed relationship between mental health and recidivism.

However, a number of foundational changes can nevertheless be made even in the absence of this additional knowledge. The fact that questions remain is not an adequate reason to reject changes to correctional management, operations, training, or treatment interventions that are already overwhelmingly supported by the evidence: A number of modifications could be developed and gainfully employed without delay.

A detailed account of every detailed aspect of each potential modification to current practice is not appropriate against the backdrop of this single program of study. However, a few general recommendations can certainly be made. These are presented briefly here. First, mental health, and once available, more comprehensive correctional profiles could be gainfully employed at various stages of case management. Such profiles provide an overview of specific areas of mental health vulnerabilities experienced by the relevant group of women, what externalizing behaviours may be expected to occur among them. Further, profiles can be used to inform case managers and treatment providers about whether a woman is likely to be ready to fully participate in and benefit from existing correctional treatment interventions aimed at reducing recidivism or whether they may require a pre-treatment acute mental health intervention to aid them in reaching a level of cognitive, emotional, and social functioning required for treatment participation. At the treatment stage, profiles and individual characteristics can be used to determine whether traditional-usually Cognitive Behaviour based- programs would suffice or if the woman should be assigned to a program with an additional emotion-focused module. Of note, emotion-focused modules should include but not necessarily be limited to include treatment targets such as emotional dysregulation, impulsivity, distress tolerance, self-care and self-

soothing techniques that can be used to avoid crises, how to build an individualized crisis management plan, and education regarding the identification of available mental health supports in the institution and in the community, as well as some methods with which the women can communicate their immediate mental health needs to these supports. For some women, it may be worth including some of their specific mental health challenges as direct treatment targets along with their traditional criminogenic needs.

Correctional agencies should take an evidence-based approach to front-line staff training. Hostility, many communication problems, risk of aggression and violence, and traumatization and stress of staff and clients can be reduced or avoided by mandatory training components. The training components should focus on gaining and understanding of, and respect for, the women's lived experiences, trauma and its effects on mental health conditions as well as externalizing behaviours. Further, mental health crisis management as well as education on how to apply dynamic as opposed to static security measures to manage both acute and ongoing problems associated with potentially disruptive acts by mentally ill clientele should be included. To reiterate, adequate education and training that empowers staff and enhances their ability to empathize and take the perspective of their most challenging clients is likely to reduce hostile attitudes, reduce power abuses and the use of punitive and illegal measures of control, increase the women's ability and willingness to consider correctional staff legitimate, trustworthy authorities, and in turn reduce the number of institutional infractions and incidences of both self-harm and interpersonal violence to a significant degree.

### **Concluding Remarks**

The current program of study constitutes a valuable, novel contribution to the existing empirical literature. The findings help inform the gender-responsive vs. gender-neutral debate

and speak to the poorly understood intersection of mental health and recidivism in the understudied population of justice-involved adult women. Current findings suggest that in this population, mental health is relevant to individual clients' risk of recidivism. The construction of mental health profiles can help guide research, foster a deeper understanding of mental health challenges among women who are in conflict with the law, and have direct practical utility across multiple correctional contexts as described directly above. The meaningful association between mental health and recidivism may help to highlight the potential utility of mental health profiles as well as clarify and emphasize the importance of understanding and developing evidence-based management and treatment strategies to correctional agencies and authorities across jurisdictions.

### References

\*Study included in the meta-analysis proper

\*\*\*Study included in the meta-analysis only in terms of background variables; no effect sizes were collected.

Hyphenated letters denote studies with overlapping samples (e.g., “ a”).

*When a study was accompanied by a separately published appendix, supplement, or annex, this document is listed directly below the main reference, indented to 1-inch.*

Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The reliability of psychiatric diagnosis revisited: The clinician’s guide to improve the reliability of psychiatric diagnosis. *Psychiatry*, 3(1), 41-50.

Aebi, M., Linhart, S., Thun-Hoenstein, L., Bessler, C., Steinhausen, H-C., & Plattner, B. (2015). *Journal of Abnormal Child Psychology*, 43, 999-1009. doi: 10.1007/s10802-01409961-y

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317-332.

Akers, R. L. (1973). *Deviant behavior: A social learning approach*. Belmont: Wadsworth.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3<sup>rd</sup> ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4<sup>th</sup> ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5<sup>th</sup> ed.). Washington, DC: American Psychiatric Association.

Andrews, D. A. (2012). The risk-need-responsivity (RNR) model of correctional assessment and treatment. In J. A. Dvoskin, J. L. Skeem, R. W. Novaco, & K. S. Douglas (Eds.),

- American psychology-law society series. *Using social science to reduce violent offending* (pp. 127-156). New York, NY: Oxford University Press.
- Andrews, D. A., & Bonta, J. (1995). *Level of Service-Inventory-Revised*. North Tonawanda, NY: Multi-health systems.
- Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*(1), 39-55. <https://doi.org/10.1037/a0018362>
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, B. (2011). Sources of variability in estimates of predictive validity A specification with level of service general risk and need. *Criminal Justice and Behavior, 38*(5), 413-432 <https://doi.org/doi:10.1177/0093854811401990>
- Andrews, D. A., Guzzo, L., Raynor, P., Rowe, B., Rettinger, L. J., Brews, A., Wormith, J.S. (2012). Are the major risk/need factors predictive of female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology, 56*(1), 113-133. <https://doi.org/10.1177/0306624X10395716>
- Baglivio, M. T., Wolff, K. T., Piquero, A. R., & Epps, N. (2015). The relationship between Adverse Childhood Experiences (ACE) and juvenile offending trajectories in a juvenile offender sample. *Journal of Criminal Justice, 43*, 229-241. <https://doi.org/10.1016/j.jcrimjus.2015.04.012>
- Baglivio, M. T., Wolff, K. T., Piquero, A. R., Greenwald, M. A., & Epps, N. (2016). Racial/ethnic disproportionality in psychiatric diagnoses and treatment in a sample of

- serious juvenile offenders. *Journal of Youth and Adolescence*, 46, 1424-1451.  
<https://doi.org/10.1007/s10964-016-0573-4>
- Baron, S. W. (1997). Risky lifestyles and the link between offending and victimization. *Studies on Crime and Crime Prevention*, 6(1), 53-71.
- Barnoski, R. P. (2003). *Evaluation of Washington State's 1996 Juvenile Court Program for High-risk, First-time Offenders: Final Report*. Olympia, WA: Washington State Institute for Public Policy
- Beaudette, J. N., Power, J., & Stewart, L. A. (2015). *National prevalence of mental disorders among incoming federally sentenced men offenders* (Report No.R-357). Ottawa, ON: Correctional Service of Canada.
- Becker, M. A., Andel, R., Boaz, T., & Constantine, R. (2011). Gender differences and risk of arrest among offenders with serious mental illness. *Journal of Behavioral Health Services and Research*, 38(1), 16-27 [https://doi.org/doi: 10.1007/s11414-010-9217-8](https://doi.org/doi:10.1007/s11414-010-9217-8)
- Beckham, J. C., & Moore, S. D. (2000). Interpersonal hostility and violence in vietnam veterans with chronic post-traumatic stress disorder: A review of theoretical models and empirical evidence. *Aggression and Violent Behaviour*, 5, 451-466.
- Belknap, J., & Holsinger, K. (2006). The gendered nature of risk factors for delinquency. *Feminist Criminology*, 1, 48-71 [https://doi.org/doi: 10.1177/1557085105282897](https://doi.org/doi:10.1177/1557085105282897)
- Benda, B. B. (2005). Gender differences in life-course theory of recidivism: A survival analysis. *International Journal of Offender Therapy and Comparative Criminology*, 49, 325-342.  
<https://doi.org/10.1177/0306624X04271194>

- Berg, M. T. (2011). Examining the neighbourhood context of the violent offending-victimization relationship: A prospective investigation. *Journal of Quantitative Criminology*, 27(4), 427-51.
- <sup>a\*</sup>Blanchette, K. (1996). *The relationship between criminal history, mental disorder, and recidivism among federally sentenced female offenders* (Unpublished Master's thesis). Carleton University: Ottawa, Ontario, Canada.
- Blanchette, K., & Brown, S. (2006). *The assessment and treatment of women offenders: An integrative approach*. Chichester, West Sussex: J. Wiley & Sons.
- <sup>a\*</sup>Blanchette, K., & Motiuk, L. (1996). Female offenders with and without major mental health problems: A comparative investigation. Research Report R-46. Ottawa, ON: Correctional Service of Canada.
- Bloom, B. E., & Covington, S. S. (2008). Addressing the mental health needs of women offenders. In R. Gido & L. Dalley (Eds.), *Women's mental health issues across the criminal justice system*. (pp. 160-176). Columbus, OH: Prentice Hall.
- Bloom, B., Owen, B., & Covington, S. (2003). *Gender-responsive strategies: Research practice and guiding principles for women offenders*. Washington, DC: USDOJ, National Institute of Corrections.
- Blume, S. (1997). Women: Clinical aspects. In J. Lowinson, P. Ruiz, A. Milkman, & J. Langrod (Eds.), *Substance abuse: A comprehensive textbook* (pp. 645-654). Baltimore: Williams and Wilkins.
- Bonta, J., & Andrews, D.A. (2017). *The psychology of criminal conduct* (6<sup>th</sup> ed.). New York: NY: Routledge Taylor & Francis Group.

- Bonta, J., Blais, J., & Wilson, H. A. (2014). A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior, 19*, 278-287. <https://doi.org/10.1016/j.avb.2014.04.014>
- Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*, 123–142. <https://doi.org/10.1037/0033-2909.123.2.123>.
- Boonmann, C., van Vugt, E. S., Jansen, L. M. C., Colins, O. F., Doreleijers, T. A. H., Stams, G.-J. J. M., & Vermeiren, R. R. J. M. (2015). Mental disorders in juveniles who sexually offended: A meta-analysis. *Aggression and Violent Behavior, 24*, 241-249. <https://doi.org/10.1016/j.avb.2015.06.003>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley.
- 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 Behavior, 39*(11), 1481-1508. <https://doi.org/doi:10.1177/0093854812456777>
- Brinded, P. M. J., Simpson, A. I. F., Laidlaw, T. M., Fairley, N., & Malcolm, F. (2001). Prevalence of psychiatric disorders in New Zealand prisons: A national study. *Australian and New Zealand Journal of Psychiatry, 35*, 166-173. <https://doi.org/10.1046/j.1440-1614.2001.00885.x>
- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

- Brown, S. (2016). *A review of the women offender risk/need research: In search of gender-neutral, women-salient, and women-specific risk factors*. Ottawa, ON: Correctional Service of Canada.
- Brown, G. P., Barker, J., McMillan, K., Norman, R., Derkzen, D., Stewart, L. A., & Wardrop, K. (2018). *Prevalence of mental disorder among federally sentenced women offenders: Incustody and intake samples* (Report No. R-420). Ottawa, ON: Correctional Service of Canada.
- Brown, S. L., Wanamaker, K. A., Greiner, L., Scott, T., & Rettinger. (2019). Complex trauma: One path to the youth justice system-A gender-informed latent profile analysis. Manuscript submitted for publication.
- Brugha, T., Singleton, N., Meltzer, H., Bebbington, P., Farrell, M., Jenkins, R., et al. (2005). Psychosis in the community and in prisons: A report from the British National Survey of Psychiatric Morbidity. *American Journal of Psychiatry*, *162*, 774-780.  
<https://doi.org/10.1176/appi.ajp.162.4.774>
- Butler, T., Allnutt, S., Cain, D., Owens, D., & Muller, C. (2005). Mental disorder in the New South Wales prisoner population. *Australian and New Zealand Journal of Psychiatry*, *39*, 407-413. <https://doi.org/10.1080/j.1440-1614.2005.01589.x>
- Butler, T., Andrews, G., Allnutt, S., Sakashita, C., Smith, N., & Basson, J. (2006). Mental disorders in Australian prisons: A comparison with a community sample. *Australian and New Zealand Journal of Psychiatry*, *40*, 272-276 <https://doi.org/10.1080/j.1440-1614.2006.01785.x>
- Brugha, T. S., Matthews, R., Morgan, Z., Hill, T., Alonso, J., & Jones, D. R. (2012). Methodology and reporting of systematic reviews and meta-analyses of observational

studies in psychiatric epidemiology: Systematic Review. *The British Journal of Psychiatry*, 200, 446-453 <https://doi.org/10.1192/bjp.bp.111.098103>

Campbell, C. A., Miller, W., Papp, J., Barnes, A.R., Onifade, E., & Anderson, V. R. (2019).

Juvenile probationers An application of latent profile analysis to a risk-need-responsivity assessment model. *Criminal Justice and Behavior*, 46(1), 82-100.

<https://doi.org/10.1177/0093854818796869>

Canadian Human Rights Commission. (April, 2012). *Submission to the Committee Against Torture*.

\*<sup>b</sup>Carroll, M. (2017). *Recidivism of women living with HIV and at-risk for HIV on probation: A recurrent events survival analysis* (Unpublished Master's thesis). Yale University: Newhaven, Connecticut.

\*\*\*Chang, Z., Larsson, H., Lichtenstein, P., & Fazel, S. (2015). Psychiatric disorders and violent reoffending: a national cohort study of convicted prisoners in Sweden. *Lancet Psychiatry*, 2, 891-900. [https://doi.org/10.1016/S2215-0366\(15\)00234-5](https://doi.org/10.1016/S2215-0366(15)00234-5)

Cheisa, M., Hanton, R., & Fontagy, P. (2011). Clinical Associations of Deliberate Self-Injury and Its Impact on the Outcome of Community-Based and Long-Term Inpatient Treatment for Personality disorder. *Psychotherapy and Psychosomatics*, 80(2), 100-109.

Chen, X. (2009). The link between juvenile offending and victimization: The influence of risky lifestyles, social bonding, and individual characteristics. *Youth Violence and Juvenile Justice*, 7(2), 119-135. <https://doi.org/10.1177/1541204008328799>

Chesney-Lind, M. (1989). Girls' crime and woman's place: Toward a feminist model of female delinquency. *Crime and Delinquency*, 35(1), 5-29.  
<https://doi.org/10.1177/0011128789035001002>

- Chesney-Lind, M. (1997). *The female offender: Girls, women, and crime*. Thousand Oaks, CA: Sage.
- Chesney-Lind, M., & Shelden, R. G. (2003). *Girls, delinquency, and juvenile justice* (3<sup>rd</sup> ed.). Belmont, CA: Wadsworth Publishing Company.
- Chesney-Lind, M., & Pasko, L. (2013). *The female offender: Girls, women and crime* (3<sup>rd</sup> ed.). California: Sage Publications.
- Chilcoat, H. D., & Breslau, N. (1998). Investigations of causal pathways between PTSD and drug use disorders. *Addictive Behaviors, 23*(6), 827-40.
- Clark, S., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. Retrieved from <http://www.statmodel.com/download/relatinglca.pdf>.
- Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology, 112*, 155-159.
- Collins, L.M., & Lanza, S. T. (2010). *Latent class and latent transition analysis With applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley & Sons, Inc.
- Coolidge, F.L., Marle, P.D., van Horn, S.A., & Segal, D.L. (2011). Clinical syndromes, personality disorders, and neurocognitive differences in male and female inmates. *Behavioral Sciences and the Law, 29*, 741-751. <https://doi.org/10.1002/bsl.997>
- Corabian, G., Appell, R., & Wormith, S. J. (2013). *Female offenders and self-harm: An overview of prevalence and evidence-based approaches*. Saskatchewan, Canada: Centre for Forensic Behavioural Science and Justice Studies, University of Saskatchewan.
- Cornelius, J. R., Kirisci, L., Reynolds, M., Clark, D. B., Hayes, J., & Tarter, R. (2010). PTSD contributes to teen and young adult cannabis use disorders. *Addictive Behaviors, 35*(2), 91-94.

- Corrado, R. R., Cohen, I., Hart, S., & Roesch, R. (2000). Comparative examination of the prevalence of mental disorders among jailed inmates in Canada and the United States. *International Journal of Law and Psychiatry*, 23(5-6), 633-647.
- Correctional Service of Canada (2009). *The changing offender population. Highlights, 2009*. Ottawa, ON: Correctional Service of Canada.
- Covington, S. (1998). The relational theory of women's psychological development: Implications for the criminal justice system. In R. T. Zaplin (Ed.), *Female offender: Critical perspectives and effective interventions* (pp.113-128). Gaithersburg, MD: Aspen Publishing.
- Covington, S. & Bloom, B. E. (2006). Gender-responsive treatment and services in correctional settings. *Women and Therapy*, 39(3-4), 9-33. [https://doi.org/10.1300/J015v29n03\\_02](https://doi.org/10.1300/J015v29n03_02)
- Cox, D. R., & Snell, D. J. (1989). *The analysis of binary data* (2<sup>nd</sup> ed.). London: Chapman & Hall.
- \*Dalbir, N. (2017). *Mental health, substance use, and co-occurring disorders among jail inmates: Prevalence, impact of recidivism, and gender differences* (Master's Thesis). Retrieved from ProQuest. (ProQuest No. 10690845)
- Daly, K. (1992). Women's pathways to felony court: Feminist theories of lawbreaking and problems of representation. *Southern California Review of Law and Women's Studies*, 2, 11-52. <https://doi.org/10.4135/9781412959193.n69>
- Davies, S., Clarke, M., Hollin, C., Duggan, C. (2007). Long-term outcomes after discharge from medium secure care: a cause for concern. *British Journal of Psychiatry*, 191, 70-74. <https://doi.org/10.1192/bjp.bp.106.029215>

- Deschenes, E. P., Owen, B., & Crow, J. (2007). Recidivism among female prisoners: Secondary analysis of the 1994 BJS recidivism data set (Doc N°. 216950). Long Beach, CA: California State University
- Dehart, D., Lynch, S., Belknao, J., Drass-Brailsford, P., & Breen, B. (2013). Life history models of female offending: The roles of serious mental illness and trauma in women's pathways to jail. *Psychology of Women Quarterly*, 38(1), 138-151.  
<https://doi.org/10.1177/0361684313494357>
- Derkzen, D., Booth, L., McConnell, A., & Taylor, K. (2012). *Mental health needs of women offenders*. Ottawa, ON: Correctional Service of Canada.
- Derzen, D., Wardrop, K., & Wanamaker, K. (February, 2019). *Women offender assessment: Can gender-informed variables improve risk prediction?* (N° R-143). Ottawa, ON: Correctional Service of Canada.
- Desmarais, S. L., & Singh, J. P. (2013). Risk assessment validated and implemented in correctional settings in the United States. Washington, DC: Council of State Governments Justice Center. [csgjusticecenter.org/wp-content/uploads/2014/07/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf](http://csgjusticecenter.org/wp-content/uploads/2014/07/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf)
- Dierkhising, C.B., Ko, S.J., Woods-Jaeger, B., Briggs, E.C., Lee, R., & Pynoos, R.S. (2013). Trauma histories among justice-involved youth: Findings from the National Child Traumatic Stress Network. *European Journal of Psychotraumatology*, 4(1), 20274.  
<http://dx.doi.org/10.3402/ejpt.v4i0.20274>.
- Donley, S., Habib, L., Jovanovic, T., Kamkwala, A., Evces, M., Egan, G., Bradley, B., & Ressler, K. K. (2012). Civilian PTSD symptoms and risk for involvement in the criminal

- justice system. *Journal of the American Academy of Psychiatry and the Law*, 40(4), 522-529.
- Du, J., Huang, D., Zhao, M., & Hesel, Y. I. (2013). Drug-abusing offenders with co-morbid mental disorders: Gender differences, problem severity, treatment participation, and recidivism. *Biomedical and Environmental Sciences*, 26(1), 32-29.  
<https://doi.org/10.3967/0895-3988.2013.01.004>
- \*Ducat, L., McEwan, T., & Ogloff, J. R. P. (2017). A comparison of psychopathology and reoffending in female and male convicted firesetters. *Law and Human Behavior*, 41(6), 588-599. <https://doi.org/10.1037/1hb0000264>
- Duncan, D. F. (1974). Drug abuse as a coping mechanism. *American Journal of Psychiatry*, 131(6), 724.
- Eher, R., Rettenberger, M., & Schilling, F. (2010). Psychiatrische Diagnosen von Sexualstraftätern: Eine empirische Untersuchung von 807 inhaftierten Kindesmissbrauchstätern und Vergewaltigern. *Zeitschrift für Sexualforschung*, 23, 23–35.  
<https://doi.org/10.1055/s-0030-1247274>.
- Ennis, A., McLeod, P., & Watt, M. C. (2014). The criminalization of the mentally ill - a Canadian, gender-based perspective. *Psynopsis*, 36(1), 11.
- \*Erickson, J. R. (2014). *Risk factors for criminal recidivism in female offenders* (Unpublished Master's thesis). Drexel University, Philadelphia.
- \*\*\*Erickson, J. R. (2016). *Relationships between abuse history, substance abuse, mental health and recidivism in female offenders* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (ProQuest No. 10252879)

- Fazel, S., Fiminska, Z., Cocks, C., & Coid, J. (2016). Patient outcomes following discharge from secure psychiatric hospitals: Systematic review and meta-analysis. *The British Journal of Psychiatry*, 208, 17-25. <https://doi.org/10.1192/bjp.bp.114.149997>
- Fazel, S., & Seewald, K. (2012). Severe mental illness in 33 588 prisoners worldwide: Systematic review and meta-regression analysis. *The British Journal of Psychiatry*, 200, 364-373 <https://doi.org/10.1192/bjp.bp.111.096370>
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . . Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*, 14(4), 245-258. [https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8)
- Field, A. (2009). *Discovering statistics using SPSS* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage Publications Inc.
- Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York, NY: Russell Sage.
- Ford, J. D., Chapman, J., Connor, D. F., & Cruise, K. R. (2012). Complex trauma and aggression in secure juvenile justice settings. *Criminal Justice and Behavior*, 39(6), 694-724. <https://doi.org/10.1177/0093854812436957>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18. <https://doi.org/10.1037/a0024338>
- \*Gehring, K. S. (2018). A direct test of pathways theory. *Feminist Criminology*, 13(2), 115-137. <https://doi.org/10.1177/1557085116646195>

- Geiser, C. (2013). *Data analysis with Mplus*. New York, NY: The Guilford Press.
- Ghaemi, S. N. (2013). Taking disease seriously in DSM. *World Psychiatry, 12*(3), 210-212.  
<https://doi.org/10.1002/wps.20082>
- Giannuli, M. M. (2001). *Mood-affective and anxiety disorders in female juvenile offenders: Prediction of crime indexes* (Unpublished doctoral dissertation). The Ohio State University, Ohio.
- Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Goodwin, J. (2020). *Exploring the relationship between reliability and predictive validity in risk assessment using the Service Planning Instrument (SPIN)*. (Unpublished master's thesis prospectus document). Carleton University.
- Green, L. (2006). *Gender influences and methodological considerations in adolescent risk-need assessment: A meta-analysis*. (Unpublished doctoral dissertation). University of Western Ontario, Ontario, Canada.
- Greenfeld, L. A., & Snell, T. L. (1999). *Bureau of Justice Statistics special report: Women Offenders*. Retrieved from <https://www.bjs.gov/content/pub/pdf/wo.pdf>
- Greiner, L. E., Law, M. A., & Brown, S. L. (2015). Using dynamic factors to predict recidivism among women A four-wave prospective study. *Criminal Justice and Behavior, 42*(5), 457-480. <https://doi.org/10.1177/0093854814553222>
- Gunter, T. D., Chibnall, J. T., Antoniak, S. K., McCormick, B., & Black, D. (2012). Relative contributions of gender and traumatic life experience to the prediction of mental disorders in a sample of incarcerated offenders. *Behavioral Sciences and the Law, 30*(5), 615-630.  
<https://doi.org/1002/bsl.2037>

- Guthrie, B. J., Cooper, S. M., Brown, C., & Metzger, I. (2012). Degrees of difference among minority female juvenile offenders' psychological functioning, risk behavior engagement, and health status: A latent profile investigation. *Journal of Health Care for the Poor and Underserved, 23*(1), 204-225. <https://doi.org/10.1353/hpu.2012.0016>
- Haider, A. H., Herrera-Escobar, J. P., Al Rafai, S. S., Harlow, A. F., Apoj, M., Nehra, D., ... Salim, A. (2018). Factors associated with long-term outcomes after injury. *Annals of Surgery, 271*(6), 1165-1173. <https://doi.org/10.1097/SLA.0000000000003101>
- Haney, C. (2002, January). The psychological impact of incarceration: Implications for post-prison adjustment [Working Paper]. *From Prison to Home Conference*, United States.
- Hannah-Moffatt, K. (2009). Gridlock or mutability: Reconsidering gender and risk assessment. *Criminology and Public Policy, 8*, 209-219. <https://doi.org/10.1111/j.1745-9133.2009.00549.x>
- Hare, R. D. (2003). *The Hare Psychopathy Checklist Revised* (2<sup>nd</sup> ed.). Toronto: Multi-health systems.
- Harpur, T. J., Hare, R. D., & Hakstian, A. R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment, 1*, 6-17.
- Harris, D. A. (1993). Factors for reasonable suspicion: When Black and poor means stopped and frisked. *Indiana Law Journal, 69*(3), 659-88.
- Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law & Human Behavior, 26*(4), 377-394. <https://doi.org/10.1023/A:101634732>

- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hedges, L. V., & Tipton, E. (2010). Chapter 58: Meta-analysis. In A. Steptoe (ed.), *Handbook of Behavioral Medicine*. IL: US: Springer Science+Business Media, LLC.
- Helmus, L., & Babchisin, K. (2013). *How to conduct a meta-analysis: A 1-day workshop: Manual*. Ottawa, ON: Carleton University.
- Herbert, M. R., Rose, J. S., Rosengard, C., Clarke, J. G., & Stein, M. D. (2007). Levels of trauma among women inmates with HIV risk and alcohol use disorders: Behavioral and emotional impacts. *Journal of Trauma and Dissociation*, 8(2), 27-46.  
[https://doi.org/10.1300/J229v08n02\\_03](https://doi.org/10.1300/J229v08n02_03)
- Herman, J. L. (1992). Complex PTSD: A syndrome in survivors of prolonged and repeated trauma. *Journal of Traumatic Stress*, 5(3), 377-391.
- Hubbard, D. J., & Pratt, T. C. (2002). A meta-analysis of the predictors of delinquency among girls. *Journal of Offender Rehabilitation*, 34(3), 1-13. doi: 10.1300/J076v34n03001
- Huebner, B. M., Dejong, C., & Cobbina, J. (2010). Women coming home: Long-term patterns of recidivism. *Justice Quarterly*, 27(2), 225-254.  
<https://doi.org/10.1080/07418820902870486>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Psychological Methods*, 3, 486-504.  
<https://doi.org/10.1136/bmj.327.7414.557>
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: Wiley.
- Houser, K., & McCord, E. S. (2018). Female offenders, mental illness and recidivism: An examination of mental illness and substance use disorders among a sample of female

- parolees released to the city of Philadelphia. In L. M. Carter, & C. D. Marcum (Eds.), *Female offenders and reentry: Pathways and barriers to returning to society* (pp. 62-75). New York, NY: Routledge.
- Hubbard, D. J., & Pratt, T. C. (2002). A meta-analysis of the predictors of delinquency among girls. *Journal of Offender Rehabilitation, 34*(3), 1-13.  
<https://doi.org/10.1300/J076v34n03001>
- Jackson, M. E., Sippel, L. M., Mota, N., Whalen, D., & Schumaker, J. A. (2015). Borderline personality disorder and related constructs as risk factors for intimate partner violence perpetration. *Aggression and Violent Behavior, 24*, 95-106 <https://doi.org/10.1016/j.avb.2015.04.015>
- Jäggi, L. J., Mezuk, B., Watkins, D. C. & Jackson, J. S. (2016). The relationship between trauma, arrest, and incarceration history among black Americans: Findings from the National Survey of American Life. *Society and Mental Health, 6*(3), 187-206.  
<https://doi.org/10.1177/2156869316641730>
- James, D. J., & Glaze, L. E. (2006). *Mental health problems of prison and jail inmates* (NCJ 213600). Washington, DC: United States Department of Justice, Bureau of Justice Statistics.
- Jennings, W. G., Piquero, A. R., & Reingle, J. M. (2012). On the overlap between victimization and offending: A review of the literature, *Aggression and Violent Behavior, 17*(1), 16-26.  
<https://doi.org/10.1016/j.avb.2011.09.003>
- Jernigan, A. S. (2000). Driving while Black: Racial profiling in America. *Law & Psychology Review, 24*, 127.

- Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2015). *Criminal Justice and Behavior*, 42(3), 321-338. <https://doi.org/10.1177/0093854814547041>
- Jones, N. J., Brown, S. L., Wanamaker, K. A., & Greiner, L. E. (2014). A quantitative exploration of gendered pathways to crime in a sample of male and female juvenile offenders. *Feminist Criminology*, 9(2), 113-136. <https://doi.org/10.1177/1557085113501850>
- Jones, N. J., & Robinson, D. (2017a). *The validation of the Service Planning Instrument (SPIn) for Alberta community corrections*. Ottawa, ON: Orbis Partners, Inc.
- Jones, N. J., & Robinson, D. (2017b). *The validity of the Service Planning Instrument for Clark County, Washington adult probation*. Ottawa, ON: Orbis Partners, Inc.
- Jones, N. J., & Robinson, D. (2018). The Service Planning Instrument (SPIn). In J. P. Singh, D. G. Kroner, S. Wormith, S. L. Desmarais, & Z. Hamilton (Eds.), *Handbook of recidivism risk/needs assessment tools* (pp. 181-198). John Wiley & Sons, Ltd.
- Joosen, K. J., Palmen, H., Kruttschnitt, C., Bijleveld, C., Dirkzwager, A., & Nieuwbeerta, P. (2016). How “gendered” are gendered pathways to prison?: A latent class analysis of the life experiences of male and female prisoners in the Netherlands. *Journal of Developmental and Life-Course Criminology*, 2, 321-340. <https://doi.org/10.1007/s40865-016-0033-665>
- Keers, R., Ullrich, S., DeStavola, B. L., & Coid, J. W. (2014). Association of violence with emergency of persecutory delusions in untreated schizophrenia. *American Journal of Psychiatry*, 171, 332-339. <https://doi.org/10.1176/appi.ajp.2013.13010134>
- Kendler, K., Aggen, S. H., Czajkowski, N., Roysamb, E., Tambs, K., Torgersen, S., et al. (2008). The structure of genetic and environmental risk factors for DSM-IV personality

- disorders. *Archives of General Psychiatry*, 65, 1438-1446.  
<https://doi.org/10.1001/archpsyc.65.12.1438>
- Kiehl, K. A., & Hoffman, M. B. (2011). The criminal psychopath: History, neuroscience, treatment, and economics. *Jurimetrics*, 51, 355-397.
- \*\*\*King, E. A., Tripodi, S. J., & Veech, C. A. (2018). The relationship between severe mental disorders and recidivism in a sample of women released from prison. *Psychiatric Quarterly*, 89, 717-731. <https://doi.org/10.1007/s11126-018-9572-9>
- Kinsler, P. J., & Saxman, A. (2007). Traumatized offenders: Don't look now, but your jail is also your mental health center. *Journal of Trauma and Dissociation*, 8(2), 81-95.  
[https://doi.org/10.1300/J229v08n02\\_06](https://doi.org/10.1300/J229v08n02_06)
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. (3<sup>rd</sup>.ed.). New York: The Guildford Press.
- Komarovskaya, I. (2009). *Trauma, PTSD, and the cycle of violence among incarcerated men and women* (Unpublished doctoral dissertation). University of Virginia, Virginia.
- Kosson, D. S., Kelly, J. C., & White, J. W. (1997). Psychopathy-related traits predict self-reported sexual aggression among college men. *Journal of Interpersonal Violence*, 12, 241-254.
- Kraska, P. B., & Kappeler, V. E. (1997). Militarizing American police: The rise and normalization of paramilitary units. *Social Problems*, 44(1), 1-18.
- \*Kubiak, S. P. (2004). The effects of PTSD on treatment adherence, drug relapse, and criminal recidivism in a sample of incarcerated men and women. *Research on Social Work Practice*, 14(6), 424-433. <https://doi.org/10.1177/1049731504265837>

Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, G. K., Marmar, C. R., & Weiss, D. S. (1990). *Trauma and the Vietnam war generation*. New York, NY: Brunner/Mazel.

Lamsma, J., & Harte, J. M. (2015). Violence in psychosis: Conceptualizing its causal relationship with risk factors. *Aggression and Violent Behavior, 24*, 75-82.  
<https://doi.org/10.1016/j.avb.2015.05.003>

Lanes, E. (2005). *Identification of variables associated with conditional release outcome for mentally disordered offenders* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global (Accession No. 3170784)

Langlois, K. A., Samokhvalov, A. V., Rehm, J., Spence, S. T., & Gorber, S. C. (2012). *Health state descriptions for Canadians: Mental illnesses* (Catalogue No. 82-619-M-No.004). Ottawa, ON: Statistics Canada, Health Analysis Division. Retrieved from the Canadian Mental Health Association website [https://www.camh.ca/-/media/files/healthstatedescriptionsforcanadians\\_mentalillnesses-eng-pdf.pdf](https://www.camh.ca/-/media/files/healthstatedescriptionsforcanadians_mentalillnesses-eng-pdf.pdf)

Lawson, B. M., Brossart, D. F., & Shefferman, L. W. (2010). Assessing gender role of partner-violent men using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2); Comparing abuser types. *Professional Psychology: Research and Practice, 41*(3), 260-266. <http://dx.doi.org/10.1037/a0019589>

Levy, M. S. (1998). A helpful way to conceptualize and understand reenactments. *The Journal of Psychotherapy Practice and Research, 7*, 227-235.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Loannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). *PLoS Medicine, 6*(7): e1000100.  
<https://doi.org/10.1371/journal.pmed.1000100>

- \*<sup>c</sup>Light, M., Grant, E., & Hopkins, K. (2013). *Gender differences in substance misuse and mental health among prisoners. Results from the Surveying Prisoner Crime Reduction (SPCR) longitudinal cohort study of prisoners*. England & Wales: Ministry of Justice Analytical Services. Retrieved from Ministry of Justice website <http://www.justice.gov.uk/publications/research-and-analysis/moj>
- \*<sup>c</sup>Ministry of Justice (2013). *Annex A Surveying Prisoner Crime Reduction (SPCR)<sup>1</sup> tables*.
- Lindgren, M., Jensen, J., Dalteg, A., Meurling, A. W., Ingvar, D. H., & Levander, S. (2002). Dyslexia and AD/HD among Swedish prison inmates. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 3, 84–95.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis. Applied social research methods series* (Vol 49). Thousand Oaks, CA: Sage.
- Liu, F., Wen, Y-F., Zhou, Y., Lei, G., Guo, Q-y., & Dang, Y-h. (2018). A meta-analysis of emotional disorders as possible risk factors for chronic periodontitis. *Medicine*, 97(28), 1-9. <https://doi.org/doi: 10.1097/MD.0000000000011434>
- Livingston, J. D. (2009). *Mental Health and Substance Use Services in Correctional Settings: A Review of Minimum Standards and Best Practices*. Vancouver, BC: International Centre for Criminal Law Reform and Criminal Justice Policy.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778 <https://doi.org/10.1093/biomet/88.3.767>
- Loeber, R., & Farrington, D. P. (2000). Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions and policy implications. *Development and Psychopathology*, 12, 737-762.

- Lord, E. A. (2008). The challenges of mentally ill female offenders in prison. *Criminal Justice and Behavior*, 35(8), 928-942. <https://doi.org/10.1177/0093854808318660>
- Skeem, J., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology*, 54(4), 680-712. <https://doi.org/10.1111/1745-9125.12123>
- Lubke, G., & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modelling*, 14, 26-47. <https://doi.org/10.1080/10705510709336735>
- Lowenstein, J., Purvis, C., & Rose, K. (2016). A systematic review on the relationship between antisocial, borderline and narcissistic personality disorder diagnostic traits and risk of violence to others in a clinical and forensic sample. *Borderline personality disorder and Emotional Dysregulation*, 3(1), 14-25. <https://doi.org/10.1186/s40479-016-0046-0>
- Magaletta, P. R., Diamond, P.M., Faust, E., Daggett, D. M., & Camp, S. D. (2009). Estimating the mental illness component of service needs in corrections. Results from a mental health prevalence project. *Criminal Justice and Behavior*, 36(3), 229-244. <https://doi.org/10.1177/0093854808330390>
- Mahoney, J. H., Jacob, J., & Hobson, H. (2017). *Women in Canada: A gender-based statistical report: Women and the criminal justice system*. (Cat. No.:89-503-X). Ottawa, ON: Statistics Canada. Retrieved from <https://www150.statcan.gc.ca/n1/pub/89-503-x/2015001/article/14785-eng.pdf>
- Mallicoat, S. L. (2011). Female offenders and their crimes. In *Women and Crime: A Text/Reader*. (pp. 343-554). Thousand Oaks, CA: Sage Publications, Inc.

- Mannerfelt, C., & Håkansson, A. (2018). Substance use, criminal recidivism, and mortality in criminal justice clients: A comparison between men and women. *Journal of Addiction, 2018*, 1-9. <https://doi.org/10.1155/2018/1689637>
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling, 16*, 191-225. <https://doi.org/10.1080/10705510902751010>
- Marshall, M. (2002). *Post-traumatic stress disorder and violence: selective processing of threat cues* (Unpublished Master's Thesis). University of Leister, England.
- McLellan, A.T., Luborsky, L., Woody, G. E., & O'Brien, C. P. (1980). An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. *The Journal of Nervous and Mental Disease, 168*(1), 26–33.
- Messina, N., Burdon, W., Hagopian, G., & Prendergast, M. (2006). Predictors of prison-based treatment outcomes: A comparison of men and women participants. *American Journal of Drug and Alcohol Abuse, 31*(1), 7-28. <https://doi.org/10.1080/00952990500328463>
- Messina, N., & Grella, C. (2006). Childhood trauma and women's health outcomes in a California prison population. *American Journal of Public Health, 96*(10), 1842-1848.
- Messina, N., & Grella, C. (2011). Who benefits from gender-responsive treatment? Accounting for abuse history on longitudinal outcomes for women in prison. *Criminal Justice and Behavior, 41*(4), 417-432. <https://doi.org/0.1177/0093854813514405>
- Messina, N., Grella, C., & Burdon, W. (2007). Childhood adverse events and current traumatic distress: A comparison of men and women drug-dependent prisoners. *Criminal Justice and Behavior, 34*(11), 1385-1401. <https://doi.org/10.1177/0093854807305150>

- McCormick, S., Peterson-Badali, M., & Skilling, T.A. (2015). Mental health and justice system involvement: A conceptual analysis of the literature. *Psychology, Public Policy, and Law*, 21(2), 213-225. <https://doi.org/10.1037/law0000033>
- McCormick, S., Peterson-Badali, M., & Skilling, T.A. (2017). The role of mental health and specific responsivity in juvenile justice rehabilitation. *Law and Human Behavior*, 41(1), 55-67. <https://doi.org/10.1037/ihb0000228>
- Miller, J. D., Dir, A., Gentile, B., Wilson, L., Pryor, L. R., & Campbell, W. K. (2010). Searching for a vulnerable dark triad: Comparing Factor 2 psychopathy, vulnerable narcissism, and borderline personality disorder. *Journal of Personality*, 78(5), 1529-1564. <https://doi.org/10.1111/j.1467-6494.2010.00660.x>
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behaviour: A developmental taxonomy. *Psychological Review*, 100, 674–701.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *British Medical Journal*, 339, 332-346. <https://doi.org/10.1136/bmj.b2535>.
- Moore, L. D., & Elkavich, A. (2008). Who's using and who's doing time: Incarceration, the war on drugs, and public health. *American Journal of Public Health*, 98(1), S176–80.
- Moore, K. E., Tull, M. T., & Gratz, K. L. (2017). Borderline personality disorder symptoms and criminal justice involvement: The roles of emotion-driven difficulties controlling impulsive behaviors and physical aggression. *Comprehensive Psychiatry*, 76, 26-35. <https://doi.org/10.1016/j.comppsy.2017.03.008>
- Morey, L. C. (1991). Personality Assessment Inventory: Professional manual. Lutz, FL: Psychological Assessment Resources.

- Muller, E., & Kempes, M. (2016). Gender differences in a Dutch forensic sample of severe violent offenders. *International Journal of Forensic Mental Health, 15*, 164-173.
- Muthén, L.K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eight Edition*. Los Angeles, CA: Muthén & Muthén.
- Nakagawa, S., & Cuthill, I. C., (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Review, 82*, 591-605.  
<https://doi.org/10.1111/j.1469-185X.2007.00027.x>
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691-692. <https://doi.org/10.1093/biomet/78.3.691>
- Neller, D., Denney, R. L., Pietz, C. A., Paul Thomlinson, R. (2006). The relationship between trauma and violence in a jail inmate sample. *Journal of Interpersonal Violence, 21*(9), 1234-1241. <https://doi.org/10.1177/0886260506290663>
- Newhill, C. E., Eack, S. M., & Mulvey, E. P. (2009). Violent behavior in borderline personality disorder. *Journal of Personality disorders, 23*, 541–554.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535-569.  
<https://doi.org/10.1080/10705510701575396>
- \*<sup>b</sup>Odio, C. D., Carroll, M., Glass, S., Bauman, A., Taxman, F. S., & Meyer, J. P. (2018). Evaluating concurrent validity of criminal justice and clinical assessments among women on probation. *Health and Justice, 6*(7), 1-10. <https://doi.org/10.1186/s40352-018-0065-6>

O'Donnell, O., House, A., & Waterman, M. (2015). The co-occurrence of aggression and self-harm: systematic literature review. *Journal of Affective Disorders, 175*, 325-350.

<https://doi.org/0.1016/j.jad.2014.12.051>

Office of the Correctional Investigator. (2019). *Annual Report 2017-2018* (Cat No. PS100).

Retrieved from <http://www.oci-bec.gc.ca/cnt/rpt/pdf/annrpt/annrpt20172018-eng.pdf>

Olivier, J., May, W. L., & Bell, M. L. (2017). Relative effect sizes for measures of risk.

*Communications in Statistics- Theory and Methods, 46*(14), 6774-6781.

<http://dx.doi.org/10.1080/03610926.2015.1134575>

<sup>d</sup>Olson, D. E., Stalans, L. J., & Escobar, G. (2016). Comparing male and female prison releasees across risk factors and postprison recidivism. *Women and Criminal Justice, 26*, 122-144

<https://doi.org/10.1080/08974454.2015.1083930>

\*<sup>d</sup>[Olson et al. unpublished univariate data]

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26*(1), 156-176. <https://doi.org/10.1037/a0035080>

Olver, M. E., & Wong, S. C. P. (2015). Short-and long-term recidivism prediction of the PCL-R and the effects of age: a 24-year follow-up. *Personality disorders, 6*(1), 97-105.

<http://dx.doi.org/10.1037/per0000095>

Orbis Partners, Inc. (2000). Youth Assessment Service Instrument (YASI). Ottawa, ON: Author.

Orbis Partners, Inc. (2003). *Service Planning Instrument (SPIn)*. Ottawa, ON: Author.

Orbis Partners. (2006). *Service Planning Instrument for Women (SPIn-W)*. Ottawa, ON: Author.

Orbis Partners, Inc. (2015). *Service Planning Instrument for Women (SPIn-W): Women's full assessment version*. Ottawa, ON: Author.

Ousey, G.C., Wilcox, P., & Fisher, B. S. (2011). Something old, something new: Revisiting competing hypotheses of the victimization-offending relationship among adolescents.

*Journal of Quantitative Criminology*, 27(1), 53-84.

\*Papadopoulos, A. K. (2011). *The prevalence and predictive nature of victimization, substance abuse & mental health on recidivism: A comparative longitudinal examination of male and female Oregon Department of Corrections inmates* (Doctoral dissertation). Retrieved from ProQuest Dissertation & Thesis Global. (UMI No. 1502147).

Parker, K. P., Lane, E. C., & Alpert, G.P. (2010). Community characteristics and police search rates. In Rice, S., White, M., (Eds.). *Race, Ethnicity, and Profiling: New and Essential Readings* (pp. 349-67). New York, NY: NYU Press

Paton, J., Crouch, W., & Camic, P. (2009). Experiences of traumatic life events: A qualitative investigation. *Clinical Child Psychology and Psychiatry*, 14(1), 43-62.

<https://doi.org/10.1177/1359104508100135>.

Pelcovitz, D., van der Kolk, B. A., Roth, S. H., Mandel, F. S., Kaplan, S. J., Resick, P. (1997). Development of a criteria set and a structured interview for disorders of extreme stress (SIDES). *Journal of Traumatic Stress*, 10(1), 3-16.

Perkins, K. R. (2010). *Examining gender differences in typological membership using latent class analysis: A novel contribution to the gender-specificity/neutrality debate* (Unpublished master's thesis). Carleton University, Ottawa, ON, Canada.

Peterson-Badali, M., McCormick, S., Vitopoulos, N., Davis, K., Haqanee, Z., & Skilling, T. A. (2015). *Canadian Criminal Law Review*, 19(1), 5-20.

- Power, J., Brown, S. L., & Usher, A. M. (2013). Prevalence and incidence of nonsuicidal self-injury among federally sentenced women in Canada. *Criminal Justice & Behavior, 40*(3), 302-320. <https://doi.org/10.1177/0093854812474427>
- Petersen, K. J., Qualter, P., & Humphrey, N. (2019). The application of Latent Class Analysis for investigating population child mental health: A systematic review. *Frontiers in Psychology, 10*, 1-16. <https://doi.org/10.3389/fpsyg.2019.01214>
- Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L., & Unnever, J. D. (2002). The relationship of attention deficit hyperactivity disorder to crime and delinquency: A meta-analysis. *International Journal of Police Science and Management, 4*, 344–360.
- Public Safety Canada (2017). *Corrections and Conditional Release Statistical Overview: Annual Report 2017* (Cat. No.: PS1-3E-PDF). Retrieved from <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ccrso-2017/ccrso-2017-en.pdf>
- Rabiner, D. L., Coie, J. D., Miller-Johnson, S., Boykin, A-S. M., & Lochman, J. E. (2005). Predicting the persistence of aggressive offending of African-American males from adolescence into young adulthood: The importance of peer relations, aggressive behaviour, and ADHD symptoms. *Journal of Emotional and Behavioral Disorders, 13*, 131–140.
- Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103-124. <https://doi.org/10.1287/mksc.12.1.103>
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior, 29*, 615–620. <https://doi.org/10.1007/s10979005-6832-7>

Rich, J. A., & Grey, C. M. (2005). Pathways to recurrent trauma among young Black men:

Traumatic stress, substance use, and the 'code of the street.' *American Journal of Public Health, 95*(5), 816–24.

Robinson, D., Van Diemen, M., & Millson B. W. (2012). The Women Offender Case

Management Model in the State of Connecticut. *Journal of Community Corrections, 21*(30), 7-25.

Rösler, M., Retz, W., Retz-Junginger, P., Hengesch, G., Schneider, M., Supprian, T.,...Thome, J.

(2004). Prevalence of attention deficit-/ hyperactivity disorder (ADHD) and comorbid disorders in young male prison inmates. *European Archives of Psychiatry and Clinical Neuroscience, 254*, 365–371.

Ross, J. M., & Babcock, J. C. (2009). Proactive and reactive violence among intimate partner

violent men diagnosed with antisocial and borderline personality disorder. *Journal of Family Violence, 24*, 607–617.

Salisbury, E. J., & Van Voorhis, P. (2009). Gendered pathways A quantitative investigation of

women probationers' paths to incarceration. *Criminal Justice and Behavior, 36*(6), 541-566. <https://doi.org/10.1177/0093854809334076>

Salisbury, E. J., Van Voorhis, P., & Spiropoulos, G. V. (2009). The predictive validity of a

gender-responsive needs assessment An exploratory study. *Crime and Delinquency, 55*(4), 550-585. <https://doi.org/10.1177/0011128707308102>

Sampson, R. J., & Laub, J. H. (1993). *Crime in the making. Pathways and turning points through*

*life*. Cambridge, MA: Harvard University Press.

- Sapers, H. (2008). *A preventable death*. Retrieved from the office of the correctional investigator's website <http://www.oci-bec.gc.ca/cnt/rpt/pdf/oth-aut/oth-aut20080620-eng.pdf?texthighlight=ashley+smith>
- Schmitt, N. (1996). Uses and abuses of Coefficient Alpha. *Psychological Assessment, 8*, 350-353.
- Schonbrun, Y. C., Johnson, J. E., Anderson, B. J., Caviness, C., & Stein, M. D. (2016). Hazardously drinking jailed women: Post-release perceived needs and risk of reincarceration. *International Journal of Offender Therapy and Comparative Criminology, 61*(16), 1819-1832. <https://doi.org/10.1177/0306624XI6634702>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333-343. <https://doi.org/10.1007/BF02294360>
- Scott, T., & Brown, S. L. (2018). Risks, strengths, gender, and recidivism among justice-involved youth: A meta-analysis. *Journal of Consulting and Clinical Psychology, 86*(11), 931-945. <https://doi.org/10.1037/ccp0000343>
- \*Scott, C. K., Grella, C. E., Dennis, M. L., & Funk, R. R. (2014). Predictors of recidivism over 3 years among substance-using women released from jail. *Criminal Justice and Behavior, 41*(11), 1257-1289. <https://doi.org/10.1177/0093854814546894>
- \*Scott, C. K., Grella, C. E., Dennis, M. L., & Funk, R. R. (2016). Time-varying model of risk for predicting recidivism among women offenders over 3 years following their release from jail. *Criminal Justice and Behavior, 43*(9), 1137-1158. <https://doi.org/10.1177/0093854816632551>

- Selenius, H., Leppänen Östman, S., & Strand, S. (2016). Self-harm as a risk factor for inpatient aggression among women admitted to forensic psychiatric care. *Nordic Journal of Psychiatry, 70*, 554-560.
- Sheffler, L. J., Stanley, I., Sachs-Ericsson, N. (2020). Chapter 4 – Aces and mental health outcomes. In Asmundson, G.J.G. & Afifi, T. O. (Eds.) *Adverse Childhood Experiences: Using evidence to Advance Research, Practice, Policy, and Prevention* (1<sup>st</sup> ed.) (pp. 47-49). London, U.K.: Academic Press. <https://doi.org/10.1016/C2017-0-03827-2>
- Skeem, J., & Lowenkamp, C. T. (2016). Risk, race, & recidivism: Predictive bias and disparate impact. *Criminology, 54*(4), 680-712. <https://doi.org/10.1111/1745-9125.12123>
- Slade, E. P., Stuart, E. A., Salkever, D. S., Karakus, M., Green, K. M., & Lalongo, N. (2008). Impacts of age of onset of substance use disorders on risk of adult incarceration among disadvantaged urban youth: A propensity score matching approach. *Drug and Alcohol Dependence, 95*(1-2), 1-13.
- Sprague, J., Javdani, S., Sadeh, N., Newman, J. P., & Verona, E. (2012). Borderline personality disorder as a female phenotypic expression of psychopathy? *Personality disorders: Theory, Research, and Treatment, 3*(2), 127-139. <https://doi.org/10.1037/a0024134>
- Song, J. W., & Chung, K. C. (2010). Observational studies: Cohort and case-control studies. *Plastic Reconstructive Surgery, 126*(6), 2234-2242. <https://doi.org/10.1097/PRS.0b013e3181f44abc>
- Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of the nonrandomized studies in meta-analyses. *European Journal of Epidemiology, 25*(9), 603-605. <https://doi.org/10.1007/s10654-010-9491-z>

- Steels, M., Roney, G., Larkin, E., Jones, P., Croudace, T., & Duggan, C. (1998). Discharged from special hospital under restrictions: a comparison of the fates of psychopaths and the mentally ill. *Criminal Behaviour and Mental Health*, 8, 39-55.
- Stewart, L.A., Wilton, G., Bablode, S., & Miller, R. (2019). *A comprehensive study of recidivism rates among Canadian federal offenders*. (N<sup>o</sup> R-426). Ottawa, ON: Correctional Service of Canada.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., ...& Thacker, S. B. (2000). *JAMA*, 283(15), 2008-2012.  
<https://doi.org/10.1001/jama.283.15.2008>
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7<sup>th</sup> ed.). NY: Pearson
- Tein, J-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling*, 20, 640-657.  
<https://doi.org/10.1080/10705511.2013.824781>
- Tripodi, S., Onifade, E., & Pettus-Davis, C. (2013). Nonfatal suicidal behavior among women prisoners: The predictive roles of childhood victimization, childhood neglect, and childhood positive support. *International Journal of Offender Therapy and Comparative Criminology*, 58(4), 394-411. <https://doi.org/10.1177/0306624X12472879>
- \*Tripodi, S., Pettus-Davis, C., Bender, K., Fitzgerald, M., Renn, T., & Kennedy, S. (2019). Pathways to recidivism for women released from prison. *Criminal Justice and Behavior*, 46(9), 1219-1236. <https://doi.org/10.1177/0093854819867119>
- Turner, K., Miller, H. A., & Henderson, C. E. (2008). Latent profile analyses of offense and personality characteristics in a sample of incarcerated female sexual offenders. *Criminal Justice and Behavior*, 35(7), 879-894. <https://doi.org/10.1177/0093854808318922>

- United Nations General Assembly. (August, 2011). Interim report of the Special Rapporteur of the Human Rights Council on torture and other cruel, inhuman or degrading treatment or punishment (A/66/268). Retrieved from United Nations website: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N11/445/70/PDF/N1144570.pdf?OpenElement>
- Usher, A., Stewart, L., Wilton, G., & Malek, A. (2010). *Profile and outcomes of male offenders with ADHD*. Ottawa, ON: Correctional Service Canada.
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*(4), 473-481. <https://doi.org/10.1037/0022-0167.51.4.473>
- van der Kolk, B. A. (2001). The assessment and treatment of Complex PTSD. In R. Yehuda (Ed.), *Traumatic Stress* (pp. 1-29). Location: American Psychiatric Press
- van der Kolk, B. A. (2005). Developmental trauma disorder: Toward a rational diagnosis for children with complex trauma histories. *Psychiatric Annals, 35*(50), 401-408.
- van der Kolk, B. A., Roth, S., Pelcovitz, D., Sunday, S., & Spinazzola, J. (2005). Disorders of Extreme Stress: The empirical foundation of a complex adaptation to trauma. *Journal of Traumatic Stress, 18*(5), 389-399. <https://doi.org/10.1002/jts.20047>
- \*Van Voorhis, P. (2012). On behalf of women offenders: Women's place in the science of evidence-based practice. *Criminology and Public Policy, 11*, 111-145. <https://doi.org/10.1111/j.1745-9133.2012.00793.x>
- \*<sup>e</sup>Van Voorhis, P., Salisbury, E., Bauman, A. & Wright, E. M. (2007). *Gender-responsive risk/needs assessment. Final Report*. [Prepared for the Maui CARE Project]. Cincinnati, OH: Centre for Criminal Justice Research. Retrieved from University of Nevada, Las

Vegas website: [https://www.unlv.edu/sites/default/files/page\\_files/27/Maui-WRNA-Final-Report.pdf](https://www.unlv.edu/sites/default/files/page_files/27/Maui-WRNA-Final-Report.pdf)

\*<sup>e</sup>Van Voorhis, P., Salisbury, E., Wright, E. M., & Bauman, A. (2008). *Achieving accurate pictures of risk and identifying gender responsive needs: Two new assessments for women offenders* (Accession No. 022844). Washington, DC, United States: U.S. Department of Justice, National Institute of Corrections. Retrieved from National Institute of Corrections website:

<https://s3.amazonaws.com/static.nicic.gov/Library/022844.pdf>

\*<sup>e</sup>Van Voorhis, P., Wright, E. M., Salisbury, E., Bauman, A., & Holsinger, K. (2009). *A gender responsive risk assessment for women offenders: Results for the Missouri Department of Corrections. Probation, prison, and pre-release samples*. Cincinnati, OH: Centre for Criminal Justice Research. Retrieved from University of Nevada, Las Vegas website:

[https://www.unlv.edu/sites/default/files/page\\_files/27/Minnesota-WRNA-Final-Report.pdf](https://www.unlv.edu/sites/default/files/page_files/27/Minnesota-WRNA-Final-Report.pdf)

\*<sup>e</sup>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 Behavior*, 37(3), 261-288.

<https://doi.org/10.1177.0093854809357442>

\*<sup>e</sup>Van Voorhis, P., Bauman, A., & Brushett, R. (2013). *Revalidation of the Women's Risk Need Assessment: Probation results. Final Report*. Cincinnati, OH: University of Cincinnati. Retrieved from the University of Nevada, Las Vegas, Centre for Crime and Justice Policy website: [https://www.unlv.edu/sites/default/files/page\\_files/27/Re-Validation-Probation-WRNA-Final-Report-2013.pdf](https://www.unlv.edu/sites/default/files/page_files/27/Re-Validation-Probation-WRNA-Final-Report-2013.pdf)

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes.

*Journal of Counseling Psychology, 51*(4), 473-481. <https://doi.org/10.1037/0022-0167.51.4.473>

Vaswani, N. (2019). The trauma, bereavement and lost experiences of women in prison.

Scotland: Centre for Youth and Criminal Justice

Verstegen, N., de Vogel, V., Huitema, A., Didden, R., & Nijman, H. (2020). Physical violence during mandatory psychiatric treatment: Prevalence and patient characteristics. *Criminal Justice and Behavior, 47*(7), 771-789. <https://doi.org/10.1177/0093854820924691>

\*Visher, C.A., & Bakken, N. W. (2014). Reentry challenges facing women with mental health problems. *Women and Mental Health, 54*(8), 768-780.

<https://doi.org/10.1080/03630242.2014.932889>

Völlm, B. A., & Dolan, M. C. (2009). Self-harm among UK prisoners: A cross-sectional study.

*The Journal of Forensic Psychiatry & Psychology, 20*, 741-751

<https://doi.org/10.1080/14789940903174030>

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17*, 228-243. <https://doi.org/10.1037/a0027127>

Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior, 27*(5), 541-558.

<https://doi.org/10.1023/A:1025490207678>

Wanamaker, K. A. (2020). *A multi-wave longitudinal examination of how strengths and risks inform risk assessment and treatment profiles for justice-involved men and women using*

- the Service Planning Instrument (SPIN)*. (Unpublished doctoral dissertation). Carleton University.
- Wanamaker, K. A., McQuaid, S., & Brown, S. L. (2016, June). *Frequencies and types of interpersonal maltreatment experienced by male and female forensic populations: A meta-analytic review*. Poster presented at the 77<sup>th</sup> Annual Canadian Psychological Association Conference, Victoria, British Columbia, Canada.
- Warren, J. I., Burnette, M., South, S. C., Chauhan, P., Bale, R., & Friend, R. (2002). Personality disorders and violence among female prison inmates. *The Journal of the American Academy of Psychiatry and the Law*, 30(4), 502-509.
- Wasserman, G. A., McReynolds, L.S., Ko, S. J., Katz, L. M., & Carpenter, J. R. (2005). Gender differences in psychiatric disorders at juvenile probation intake. *American Journal of Public Health*, 95(1), 131-137. <https://doi.org/10.2105/AJPH.2003.024737>
- Welch, C. L. (2007). *Psychosocial profiles: Serious and chronic female juvenile offenders with and without a substance use disorder* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.
- Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., ..et al (2000). *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-analyses*. Ottawa, Ontario: Ottawa Hospital Research Institute. Retrieved from [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp)
- Wibbelink, C. J. M., Hoeve, M., Stams, G. J. J. M., & Oort, F. J. (2017). A meta-analysis of the association between mental disorders and juvenile recidivism. *Aggression and Violent Behavior*, 33, 78-90. <https://doi.org/10.1016/j.avb.2017.01.005>

- Wilson, D. B. (2001). *Practical meta-analysis effect size calculator* [Online calculator].  
<https://campbellcollaboration.org/research-resources/effect-size-calculator.html>
- Wilson, D. B. (2005, May). Meta-analysis macros for SAS, SPSS, and Stata. Retrieved, February 2nd, 2019, from <http://mason.gmu.edu/~dwilsonb/ma.html>
- Wittebrood, K., & Nieuwbeerta, P. (1999). The link between offending, lifestyle and violent victimization. *European Journal on Criminal Policy and Research*, 7(1), 63-80.
- Wright, E. M., Salisbury, E. J., & Van Voorhis, P. (2007). Predicting the prison misconducts of women offenders The importance of gender-responsive needs. *Journal of Contemporary Criminal Justice*, 23(4), 310-340. <https://doi.org/10.1177/1043986207309595>
- 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 Behavior*, 39(12), 1612-1632. <https://doi.org/10.1177/0093854812451088>
- Zilberstein, K. (2014). Neurocognitive considerations in the treatment of attachment and complex trauma in children. *Clinical Child Psychology and Psychiatry*, 19(3), 336-354.  
<https://doi.org/10.1177/1359104513486998>

## Appendix A

### Non-English Sources and Search Terms

A number of additional non-English sources (Norwegian, Swedish, Danish, and German) were searched. Both English and the non-English search terms used to search foreign sources because some publications were also published in English on these sites.

#### Norway

- The Norwegian University of Science and Technology (NTNU)
- The University of Oslo (UIO; including the PubPsyc database)
- Current Research Information System in Norway (CRISTIN)
- Helsebiblioteket (a Norwegian library of health-related studies)
- Kriminalomsorgen (i.e., the department responsible for corrections in Norway)

*Norwegian search terms:* Search terms defined the gender *kvinner* (women), the justice-involved population, *forbrytere* (criminals), *gjerningsmenn* (perpetrators), *dömte* (convicted), *innsatte* (inmates/detainees), *fanger* (prisoners) *lovbrytere* (individuals who break the law), *kriminelle* (criminals), general terms for mental disorders *mental\** (mental), *mental lidelse\** (mental disorder), *mental sykdom\** (mental illness), *psykiatrisk sykdom\** (psychiatric illness), *psykisk lidelse\** (psychiatric disorder), *sinnslidelse\** (a disorder of the mind/spirit), *sinnsykdom\** (an illness of the mind/spirit), *mental helse* (mental health), *psykiatrisk\** (psychiatric), and finally, the outcome *tilbakefall* (recidivism).

#### Sweden

- Kriminalvården (i.e., the department responsible for corrections in Sweden)

*Swedish search terms:* Search terms defined the population (*kvinnor* (women), *kvinnor* (female), *kvinnliga gärningsmän\** (justice-involved women), *dömda kvinnor* (convicted women), *häktade* (inmates/detainees), *lagbrytare* (offenders), *brottslingar* (criminals), general terms for mental disorders (*mental\** (mental), *mental hälsa* (mental health), *mental sjukdom* (mental disorder), *psykiatrisk\** (psychiatric), *psykiatrisk störning\** (psychiatric disorder), *psykiatrisk sjukdom* (psychiatric illness), and finally, the outcome *äterfall* (recidivism).

### **Denmark**

- Kriminalforsorgen (i.e., the department responsible for corrections in Denmark),

*Danish search terms:* Search terms defined the population *kvinder\** (women\*), (*kvindelige lovovertrædere* (women who break the law), *gerningsmand* (perpetrator), *indsatt\** (inmate\*), *fange\** (prisoner\*), general terms for mental disorders *mentalt helbred* (mental health), *psykisk sygdom* (mental illness), *psykisk lidelse* (mental disorder), *psykiatrisk\** (psychiatric), and finally, the outcome *tilbagefald* (recidivism), *gentagelsestilfælde* (reoffending), and *recidivist* (recidivist).

### **Germany**

- Kriminologische Zentralstelle (KRIMZ; a German database for research in criminology)
- Bundesministerium der Justiz und für Verbraucherschutz (Federal ministry of justice and consumer protection),

*German search terms:* Search terms defined the population (*täterinnen* (justice-involved women), *weiblicher täter* (female offenders), *täter* (offender), *haftung\** (prisoner), *insasse\** (inmate), general terms for mental disorders *psychische gesundheit* (mental health), *psychische störung* (mental disorder), *geisteskrankheit* (mental/spirit sickness/illness), *psychiatrisch\**

(psychiatric), and finally, the outcome *rückfälligkeit* (recidivism), *wiederbeleidigen* (reoffending), *rückfälliger* (recidivist), and *wiederholungstäter* (criminal recidivist).

### **Conclusion**

Of note, thorough searches of the non-English sources did not result in any additional studies being included in the meta-analysis. A single Norwegian study to date has collected data on both mental health and recidivism among adult women (*Norwegian Offender Mental Health and Addiction Study*) but upon contacting the authors it became clear that their study was yet in protocol form, as ethics approval to start analyses had yet to be confirmed (Personal communication with Dr. Anne Bukten at <https://www.researchgate.net> September 6<sup>th</sup>, 2019).

## Appendix B

### Meta-analysis Coding Protocol

#### What to Code and What NOT to Code

##### Samples

The studies to be included in this meta-analysis can vary in terms of samples types.

##### *Accepted Sample Types*

- Women legally declared mentally disordered by the courts (MDO)\*
- Women found not criminally responsible on account of mental disorder (NGRI)\*
- Mixed single/multiple samples of disordered and non-disordered justice-involved women
- Women at all stages of criminal-justice proceedings, including charged pre-trial detainees.

*We do not accept samples of mentally disordered women without a criminal record who are followed up to assess post-release (i.e., from non-forensic psychiatric facilities) violence.*

*\*Following coding completion of all originally included studies, it was determined that samples of MDOs should be excluded. No NGRI samples were identified.*

##### **Predictors: Mental Health Variables**

We code the vast majority of mental health diagnoses and indicators, with four main exceptions; substance abuse/use/dependence disorders, APD, psychopathy, and mixed, potentially internally contradictory mental health indices enveloping numerous indices of mental disorder and mental health status in a way that prevents evaluation of any one of the indices (e.g., the Emotional/Personal Domain of the LSI-R risk assessment tool).

##### **Types of Analyses**

We code data from both univariate and multivariate analyses. Although effect sizes from multivariate analyses cannot be included in the quantitative synthesis, coding the multivariate data allows for a precise and comprehensive systematic review of the literature. In other words, by coding the data pertaining to studies that reported (only) multivariate effects, we can more easily provide a good summary of how the studies in this area have been conducted and what their findings are.

##### **Outcome Variables: Recidivism**

We code data pertaining to all recidivism outcomes (e.g., general, violent, and sexual recidivism; rearrests, reconvictions, etc). We code data pertaining to analyses in which recidivism was measured dichotomously as well as continuously (i.e., hazard ratios [HRs] from cox regressions examining time-to-recidivism). Note that we do not accept instances of violent

behaviour while institutionalized or while in custody as recidivism events, nor do we accept first-time instances of criminal or violent behaviours among patients released from secure mental health units unless the patients had previous criminal offenses on record.

### Study inclusion criteria

- **Longitudinal design** (retrospective or prospective)
- **A sample of justice-involved adult women ( $N \geq 10$ ).**
- **A measure of a mental health problem** (e.g., a specific mental disorder)
- **A outcome measure of recidivism**
- **Report sufficient information to calculate an effect size.**
- **The study must be available in English, German, Norwegian, Danish, or Swedish**

*Of note, at the end of the coding process, the original inclusion criteria were further refined, such that 1) recidivism outcomes had to be binary, 2) effect sizes had to originate from univariate rather than multivariate analyses (i.e., hazard ratios were excluded), and that studies of MDOs or NRGI women originally collected were to be excluded.*

---

### Meta-analysis Coding Protocol Instructions

This manual provides information about study coding procedures. The manual consists of four main sections: (1) study characteristics (Table 1), (2) sample characteristics (Tables 2), (3) effect size data for a) group comparisons (Table 3) and b) correlational data (Table 4). The coding manual proper is preceded by a set of general coding instructions and tips to help ease the coding process.

#### General Instructions

- If the required information is not available, enter “x” in the relevant table cell. If the information does not pertain to your effect size data enter “n/a” in that cell.
- If there are multiple studies with partially or completely overlapping samples that contain desirable data, code them all using the same coding sheets. Provide both references in Section 1 and note the overlap in NOTES in Table 1. In Tables 2, 3, and 4, indicate which study/document and page number the data came from.
- If there is more than one effect size available (i.e., either a group comparison [Table 3] or a correlation [Table 4]), code ALL available effect sizes.

- Code all numbers to three decimals.

### Group labelling

In all sections and tables, provide each sample with a group label to distinguish them from one another and to ease linking of data in Tables 1 and 2 to data in Tables 3 and 4. Use intuitively reasonable labels. If the sample is special, such as MDO or NGRI, note that in the label. Indicate if group is before or after attrition at follow-up, if required for the specific study. If the group consists of all the disordered or non disordered women (group comparison studies), an example label might be the title of their disorder (e.g., ADHD and non-ADHD). Whenever possible, also indicate if the sample was recruited from a currently active treatment program for any mental health issue or for substance abuse. Treat these two treatments as separate (TREATMH vs. TREATSA). The principle is just to be clear about the specific group to which you are referring, to whatever level of detail necessary.

### Table 1B: Study Characteristics

- **Study reference:** Reference for the study in full APA format.
- **Coder's first name:** Your name.
- **Citation:** Study citation, including year.
- **Status:** Published or unpublished.
- **Record Type:** Type of document/study (e.g., *journal article, dissertation, etc.*).
- **Peer-review:** Yes or no.
- **Location:** Study's country of origin.
- **Study design:** Prospective or retrospective.
- **# of unique ES coded from the Study/Studies:** After coding Tables 3 and 4, count the number of non-overlapping, independent effect sizes coded for the study/studies. If there were more than one documents/studies pertaining to the same sample, report the total number of unique ES coded from all relevant documents.
- **Recidivism Base Rate:** Enter the base rate for recidivism in the study for all recidivism outcome measures used. To calculate the base rate, divide the number of recidivists by the total sample size ( $N_{\text{recidivists}}/N_{\text{total}}$ ). In the event that there are multiple recidivism outcome measures (e.g., rearrests and reincarcerations), please enter recidivism base rates for all of them when possible. If a study contains multiple independent, non-overlapping samples, report base rates for each sample as well as for the combined total sample.

- **Overlap:** Indicate whether the relevant sample(s) overlap with samples from other studies that you have or will be coding. If the overlap is not discovered until the related study is being coded, this cell must be subsequently be updated in the original coding sheets. If there is overlap, provide the full reference (APA) for the related study.
- **NOTES:** If there is information that could be pertinent to the coding of the study that could not be provided in the existing tables, note them here.

## Table 2B. Sample Characteristics

### General Instructions

- In Table 2 we code the characteristics of the original total sample (before attrition), the total sample with follow-up data (i.e., after attrition), and all the subsamples of women with and without a mental health indicator, problem, or diagnosis reflected in codeable effect size data. For example, a study might report characteristics of the total sample of initially recruited women, of the total sample with follow-up data, and of women with and without a diagnosis of major depression. In this case, you would report sample characteristics of all four groups.
- If effect size data is reported for a group for which no sample characteristics are available, make a note below Table 2 to explain why no data is present for that group.
- **Important!** Report only characteristics of samples that have corresponding effect size data to be coded in Tables 3 and/or 4. If there is a subsample without relevant effect size/recidivism outcome data, do not code it in Tables 2.

### Variables

- **Page #:** Enter the page number on which the data were reported. If multiple pages, include all. If the data were reported more than once, but inconsistently, (e.g., page 13 reports 40% and while page 14 reports 40.2%), indicate which page number(s) you chose and why in a table note.
- **Overall Sample before Attrition (*N*):** Group label and sample size before attrition.
- **Overall Sample after attrition (*N*):** Group label and sample size after attrition.
- **Group 1 (*n*).** Subgroup 1 label and subsample size.
- **Group 2 (*n*).** Subgroup (e.g., comparison group) 2 label and subsample size.
- **Important!** In some studies the authors may report the total sample size but, because of missing data, report sample characteristics only for a certain proportion of that total sample (e.g., the overall sample size to be coded might be  $N = 300$ , but data on the “mean age” variable may be based on  $N = 298$  offenders (i.e., age was unknown for  $n = 2$ )).

women). When this occurs, code all variables as you normally would, but enter the reduced sample size in parentheses in the cells for the relevant variables.

**Predominant Disorder:** Enter the most common mental disorder for the sample/subgroup and the percentage of the sample/subgroup with this disorder. This should only be coded for binary diagnoses and not for other types of mental health indicators.

**Disorders:** Enter the name of each disorder along with the percentage of women diagnoses with it. Repeat for all disorder to the level of detail provided in the study. This variable is only to be coded for binary diagnoses.

*Of note, when available, also provide the percentage of women with APD and/or Psychopathy.*

- **Mean age (SD):** Enter the average age and standard deviation.
- **Average Risk Level:** Enter the average risk level of the sample(s).
- **Majority Risk Level ( $\geq 60\%$  or *mixed*):** Enter risk level for the majority ( $\geq 60\%$ ) of the sample. If no risk level is shared by 60% or more of the sample, enter “*mixed*”.
- **Risk Level (*low, low-mod, mod, mod-high, and high*):** Enter the percentage of women classified as with *low, low-moderate, moderate, moderate-high, and high risk* to reoffend. Enter whatever risk classifications or bins provided in the study.
- **Risk Type Assessed:** Enter whether they estimated risk of general, violent, or sexual recidivism.
- **Risk Assessment Type:** Enter the type of risk assessment used to estimate risk; *actuarial, structured professional judgment (SPJ), unstructured clinical judgment, or other*. If *other*, please specify.
- **Name of Risk Assessment Instrument:** Enter the name of the risk assessment instrument.
- **Index Offense:** Enter the percentage of women with non-violent only, violent, and sexual index offenses.
- **Majority Race:** If one race made up 60% or more of the sample, enter that race (e.g., *Caucasian*), along with the exact corresponding percentage. Otherwise enter *mixed*.
- **Race:** Enter the percentage of women of all reported races. Utilize the labels the original study authors use.
- **Setting:** Enter the setting of the mental health assessment; state/federal prison, remand/local jail, mental health (forensic) hospital/secure unit, or community.

Importantly, this variable is about the timing of the assessment itself, rather than, for example, the setting at initial recruitment.

- **Treatment/Medication Compliant:** Enter the percentage of women who were a) receiving mental health treatment currently and b) who were medication compliant. This pertains to mental health treatment only, excluding substance abuse.
- **Treatment:** Indicate what percentage of women received treatment just prior to, during incarceration or supervision period, or while at risk during follow-up. Specify when treatment was received; *prior to index offense, during incarceration/supervision period for index offense, or while on conditional release/during follow-up*. Treatments received prior to sentencing for index offense should not be coded here.
- **Attrition at Follow-up:** Enter the percentage of the sample excluded from recidivism analyses due to a lack of follow-up data.

**Important!** Report the attrition rate indicated by the authors. Do not rely on the reported size of the sample involved in analyses, because a reduced number of cases involved in analyses could be attributable to causes other than attrition at follow-up, such as missing data on the predictor variable.

*Note that the original coding protocol, including that submitted for registration in PROSPERO initially included two tables of risk of bias variables. Refer to Appendix C for a summary of findings for this variables and a full account of the reason for their exclusion.*

### Tables 3B and 4B: General Instructions and Overview

- When analyses involved group comparisons or simple proportions use Table 3. When the effect size data is correlational, use Table 4.
- Each table details effect size data for one single mental health predictor. If a study reports on multiple mental health predictors, copy the relevant table (both 3 and 4) as many times as required and code all variables in these tables as well.

*Note that for Table 3, in the case that multiple group comparisons are made reflecting identical contextual variables (i.e., same predictor, same outcome), you can chose to add more rows in the existing table and report the effect size data there rather than copying table, which requires re-entry of all data.*

- When multiple group comparisons are available, code them all in Table 3. Provide effect size data for all possible comparisons, regardless of whether it was specifically reported by the authors. For example, if authors provide an effect size only for one main group comparison, but the necessary proportions for additional comparisons are extractable, code all possibilities. This principle is limited to comparisons that could plausibly be of interest in the meta-analysis however. For example, comparing ADHD and non-ADHD groups is sensible. Comparing the ADHD group to the BPD group, is not.

- If several effect size measures are reported for the same analysis (e.g., Cohen's *d* and Odds Ratios (*ORs*)), report all of them. Using the split cell function eases this task.

### Tables 3B and 4B: Coding the Contextual Variables

- **Citation, Page Number, and Table Number:** Citation, page- and table number from which data were extracted.

- **(Mental health) Diagnosis/Indicator:** Name of mental health predictor.

This can be a general indicator such as “*psychiatric hospitalizations*” or a specific disorder such as “*major depression*”. If authors specify an exact diagnosis (e.g., major depression), code it “*major depression*” rather than just “*depression*”, as the latter may include other related mood disorders (e.g., dysthymia).

**Important!** If the mental health indicator is *any mental disorder*, then you must enter a table note specifying all the disorders included in the definition of any mental disorder in the original study. The same applies for the indicator *severe/major mental disorder/illness*.

- **Dichotomous or Continuous Measurement:** Measurement of the mental health predictor; *Binary, categorical, or continuous*.
- **Current or Lifetime:** Lifetime or current diagnosis/indicator.
- **Assessment Method:** Method of mental health assessment; *self-report, clinical interview, file review, correctional case management/risk/need/intake assessment, or researchers interview*.
- **Protocol:** If the assessment method was a clinical interview, enter the protocol used to structure the interview and determine the diagnosis.
- **Name of Assessment Tool:** Name of the measure used to assess the disorder. In [brackets] provide the abbreviated title of the measure, followed by the original citation (author, year) for the tool here. For non-diagnostic measurement, enter *N/A*.
- **Recidivism Type:** Recidivism outcome type; *general, violent, or sexual*. When reported, specify whether violence included sexual offenses or not. Further, if additional specifications of the outcome were made (e.g., fire-related offenses only; violent offenses excluding some subset, etc)
- **Recidivism Definition:** Recidivism definition; *re-arrests (with/without breaches), revocations, new charges (with/without technical breaches), re-convictions, re-*

*admissions to secure unit (including or excluding technical violations), re-incarcerations (with/without technical breaches), technical violations/breaches, or self-reported recidivism.* If authors reported a unique definition such as *Any fail*, code it using their selected term but add a table note explaining its exact meaning.

- **Recidivism Source:** Source of recidivism information: *official records vs. non-official sources.*
- **Recidivism Measurement:** Dichotomous or continuous recidivism measurement.
- **Length of follow-up:** Length of time sample was followed at risk in the community following release before recidivism data was collected. If study design was retrospective, simply report the time elapsed between assessment and end of follow-up. An absolute follow-up period is preferable, but reporting the average length is acceptable when the duration of follow-up was variable across women. Report duration in months.
- **Analysis Type:** Origins of effect size; *univariate or multivariate analyses.* For multivariate analysis, enter a table note specifying all the variables included in the analyses, regardless of entry order or method.

#### **Exclusive Table 3B Variables (Not applicable to Table 4B)**

- **LINK #:** It is possible that some studies will include more than one measure of recidivism for the same mental health indicator. For this reason, even for non-overlapping groups, there needs to be an indication of whether the effect size in question is independent from all other effect sizes coded from the same study. To do this, enter the same number for non-independent effect sizes (e.g., 1, 1) and different numbers for independent effect sizes (1, 2). **Non-independence;** effect sizes that involve partly or completely overlapping samples across groups and/or measures. If some or all of the same participants are involved in different effect sizes, they are non-independent even if they use different measures or some different subgroups of participants.
- **Group 1 and Group 2 Labels:** Group labels matching those used in Table 2.
- **Independent (IND):** Report independence or dependence of the subgroup from any larger sample (if valid effect size data were reported for this larger sample); *yes vs. no.*
- **n:** subsample size of both groups.
- **Recidivism:** Number and/or percentage of women in each group who recidivated.

#### **Exclusive Table 4B Variables (Not applicable to Table 3B)**

- **Subgroup non-independence.** If the reported effect size data pertains to a subgroup, indicate whether this subsample overlaps with a overall, larger study sample; *yes vs. no.*

- **N:** Total sample size involved in the effect size estimate/correlation.
- **Recidivists and non-recidivists: *n*, *Mean*, *SD*:** Enter number and corresponding proportions of recidivists and non-recidivists, as well as their respective mean scores and standard deviations on the mental health assessment measure. Enter “x” in any cell for which data is not available.

#### Effect Size Data: Tables 3B and 4B

- **Effect Size Direction:** If a positive effect size indicates that the group **with** the mental health predictor had higher recidivism rates, enter *positive*. If a positive effect size indicates that the subgroup **without** the mental health problem had higher recidivism rates, enter *negative*.
- **Effect Size Type:** Name the effect size measure. (e.g., *OR*). If multiple are reported, code all in separate (split) rows.
- **Effect Size, CI, *p*-values/levels, & Standard Errors (SE):** Numeric value of the effect size, followed by 95% confidence intervals, the *p*-value/level, as well as standard errors where applicable.

*A note on selecting a Table. Sometimes the reported effects size can be conceptualized as both group comparisons and as correlations. For example, an OR from an analysis involving a binary recidivism outcome can be understood as the difference in odds between groups (comparative) or as a simple coefficient akin to a Pearson's *r* correlation. The main goal is simply to accurately denote all relevant data available as precisely as possible. Thus, select the table that allows for the most precise rendition of all factors involved in the effect size. If this too remains unclear, rely on original study authors' reporting of the effect and select a table accordingly. Recall that unresolved issues can be noted in Section 1, Table 1, under NOTES.*

The coding sheets follow on the next page below.

**Table 1B.**

*Study Characteristics*

<b>Study reference (full APA):</b>	
<b>Coder's first name</b>	
<b>Citation (including year)</b>	
<b>Status (<i>Published/Unpublished</i>)</b>	
<b>Peer reviewed (<i>yes/no</i>)</b>	
<b>Record type (e.g., <i>dissertation</i>)</b>	
<b>Study design (<i>pro- vs. retrospective</i>)</b>	
<b>Location (country of origin)</b>	
<b># Unique ES coded</b>	
<b>Recidivism base rate(s)</b>	
<b>Overlap with group(s) from another study (<i>yes/no</i>)</b>	
<b>If overlap, a full reference for the other study is entered here. Otherwise <i>n/a</i></b>	
<b>NOTES:</b>	

**Table 2B.**

*Sample Characteristics*

<b>Sample characteristics</b>	<b>Responses</b>	<b>Page #</b>
<b>Sample Label and Size (before attrition) (N)</b>		
Sample definition		
Predominant disorder (% diagnosed)		
Disorders (% diagnosed for all disorders)		
Mean age (SD):		
Average risk level		
Majority risk level (≥60% or “mixed”)		
Risk level (% classified in each category or bin)		
Risk type assessed (general, violent, or sexual)		
Type of risk assessment (e.g., Actuarial, SPJ)		
Name of risk assessment tool		
Index offense (%non-violent, violent, and sexual)		
Majority race (≥ 60% and % or “mixed”):		
Race representation (% of sample of all reported races)		
Setting (e.g., prison, forensic MH hospital, community)		
Current treatment & medication compliance (% both)_		
MH received (% & timing)		
Attrition at follow-up (%)		
<b>Sample Label and Size (after attrition) (N)</b>		
Sample definition		
Predominant disorder (% diagnosed)		
Disorders (% diagnosed for all disorders)		
Mean age (SD):		
Average risk level		
Majority risk level (≥60% or “mixed”)		
Risk level (% classified in each category or bin)		
Risk type assessed (general, violent, or sexual)		
Type of risk assessment (e.g., Actuarial, SPJ)		
Name of risk assessment tool		
Index offense (%non-violent, violent, and sexual)		
Majority race (≥ 60% and % or “mixed”):		
Race representation (% of sample of all reported races)		
Setting (e.g., prison, forensic MH hospital, community)		
Current treatment & medication compliance (% both)_		
MH received (% & timing)		

**Table 2B.**

*Sample Characteristics*

<b>Subgroup characteristics</b>	<b>Responses</b>	<b>Page #</b>
<b>Subgroup 1 Label and Size (N)</b>		
Sample definition		
Predominant disorder (% diagnosed)		
Disorders (% diagnosed for all disorders)		
Mean age (SD):		
Average risk level		
Majority risk level (≥60% or “mixed”)		
Risk level (% classified in each category or bin)		
Risk type assessed (general, violent, or sexual)		
Type of risk assessment (e.g., Actuarial, SPJ)		
Name of risk assessment tool		
Index offense (%non-violent, violent, and sexual)		
Majority race (≥ 60% and % or “mixed”):		
Race representation (% of sample of all reported races)		
Setting (e.g., prison, forensic MH hospital, community)		
Current treatment & medication compliance (% both)_		
MH received (% & timing)		
Attrition at follow-up (%)		
<b>Subgroup 2 Label and Size (N)</b>		
Sample definition		
Predominant disorder (% diagnosed)		
Disorders (% diagnosed for all disorders)		
Mean age (SD):		
Average risk level		
Majority risk level (≥60% or “mixed”)		
Risk level (% classified in each category or bin)		
Risk type assessed (general, violent, or sexual)		
Type of risk assessment (e.g., Actuarial, SPJ)		
Name of risk assessment tool		
Index offense (%non-violent, violent, and sexual)		
Majority race (≥ 60% and % or “mixed”):		
Race representation (% of sample of all reported races)		
Setting (e.g., prison, forensic MH hospital, community)		
Current treatment & medication compliance (% both)_		
MH received (% & timing)		
Attrition at follow-up (%)		

**\*COPY THIS TABLE AS MANY TIMES AS REQUIRED TO REPORT CHARACTERISTICS OF ALL RELEVANT (SUB) GROUP.**





## Appendix C

### Risk of Bias in Observational Research

*Risk of bias* variables should not be confused with overall *study quality ratings*.

Favorable bias scores can co-occur simultaneously with low study quality ratings.

*Risk of bias variables primarily attempt to assess the degree of confidence with which the study findings' representativeness and generalizability, meaning the degree to which they can reasonably be assumed to apply to the general population of justice-involved women.*

Importantly, a number of risk of bias variables were developed separately for cohort and case-control designs specifically for the current meta-analysis. Only cohort designs were identified for inclusion, so the case-control items were never attempted coded.

In their review, Stroup and colleagues (2000) developed a very small number of guidelines for reporting of meta-analyses of observational studies in epidemiology (Stroup et al., 2000). These guidelines were based on a review of previously published meta-analyses that resembled the current meta-analysis. The authors noted that 1) whenever possible; an attempt should be made to assess risk of bias, using a combination of the most relevant existing guidelines and standards, including observational studies, 2) those items clearly relevant to the individual study and its idiosyncrasies should be the ones selected – use criteria that maximizes one's ability to assess bias in a study as it is – rather than strict and narrow criteria provided as blanket solutions for all assessments, and 3) its preferable to rely on single item scores than poorly weighted and summed total scores. These recommendations were utilized in the current study, and three subsequent sources of risk of bias items were utilized to create a five-item list for cohort designs: 1) the section of PRISMA reporting standards for meta-analyses that focus on addressing risk of bias in individual studies as well as in the eventual outcomes of the meta-

analysis (Liberati et al., 2009; Moher et al., 2009), Wells and Colleagues' (2000) Newcastle-Ottawa Scale (NOS; see Fazel et al., 2016; Liu et al., 2018 for examples of how the NOS has been utilized so far), and Song and Chung's (2010) article, which outlines study designs and methodology for observational studies were used in tandem to isolate potentially viable risk of bias items. Note that Song and Chung (2010) contributed a single item; *whether the attrition at follow-up exceeded 20%*. The authors reported that, as a rule of thumb, attrition at follow-up should not exceed 20% (Song & Chung, 2010)- attrition beyond 20% is likely to result in incomparable, no longer representative groups from the population of interest.

For cohort studies, the following variables were coded: 1) the representativeness of the a) overall sample, b) the disordered cohort(s), c) and the disordered cohort(s), 2) cohort recruitment (i.e., whether the disordered and non-disordered cohorts were recruited from the same underlying population), and lastly, 4) whether the overall attrition rate at follow-up exceeded 20%.

For case-control studies the following items were initially selected. First, whether was adequacy of definitions of both cases and controls (*yes vs. no*) were established should be determined. Second, the presence of selection bias (*yes vs. no*) must be identified. Third, whether either a-priori or at least post-recruitment group equivalency of cases and controls was ensured should be assessed (*yes vs. no*). Fourth, whether confounding variables had been controlled for (i.e., overall and for the following individual confounds deemed relevant to the current meta-analysis; Cluster B Personality disorders (e.g., BPD), Psychopathy, Substance Dependence, Risk, Age, Ethnicity, Security Classifications, and Index Offense Type; *yes vs. no*) must be established. Finally, whether equivalent assessment methods were used across cases and controls (*yes vs. no*) must be determined.

**Conclusion**

As explained previously, there were a multitude of reasons why risk of bias, although technically coded for all studies, was not utilized in analyses or reported on further in the current study. A large number of problems lead to the abandonment of risk of bias analyses. First, unacceptable difficulties in determining the accurate code for each item in each individual study, which was partly due to the lack of adequate reporting but also occurred on the account of simple definitional problems (e.g., what population did the study mean to sample from?) was a major barrier to successful utilization of items. Second, considerable inconsistencies in terms of how the same item was coded across studies were uncovered without a method with which to avoid this being identified. Thirds, the items' utter lack of utility in analyses (e.g., complete lack of variability of scores, difficulty assessing how they might be compared across studies, impossibility of determining how findings could be summarized into a coherent whole and how the individual scores may or may not affect the reliability and validity of the current meta-analytic findings).

Of note, the next page presents a simple summary table of the individual study findings for risk of bias variables.

**Table 1C.**

*Risk of Bias Variables by Individual Study*

Study name	Representativeness			Sample selection	Overall attrition > 20%
	Disordered Cohort	Non-disordered cohort	Overall sample		
Blanchette (1996); Blanchette & Motiuk (1996)	No	Yes	Yes	Yes	No
Olson et al. (2016)	No	Undetermined	Undetermined	Yes	N/A
Papadoplous (2011)	No	Yes	Yes	Yes	N/A
Scott et al. (2014; 2016)	No	No	No	Yes	No
Mannerfelt & Håkansson (2018)	No	No	No	Yes	Yes
Tripodi et al. (2019)	Yes	Yes	Yes	Yes	No
Visher & Bakken (2014)	No	No	No	Yes	Yes
Ducat et al. (2017)	Yes	Yes	Yes	Yes	Yes
King et al. (2018)	No	Yes	Yes	Yes	No
Gehring (2011)	No	No	No	Yes	N/A
Kubiak (2004)	No	No	No	Yes	Yes
Van Voorhis et al. (2007-2010)					
Maui Probation Sample	Undetermined	Undetermined	Undetermined	Yes	No
Missouri Probation Sample	Undetermined	Undetermined	Undetermined	Yes	No
Missouri Pre-release Sample	No	No	No	Yes	No
Missouri Prison Sample	No	No	No	Yes	No
Light et al. (2013)	No	No	No	Yes	No
Erickson (2014; 2016*)	No	No	No	Yes	No
Van Voorhis et al. (2012; Total sample)	No	No	No	Yes	No
Van Voorhis et al. (2013; Total sample)	No	No	No	Yes	No

*Note.* Sample selection = whether the disordered and non-disordered cohorts were recruited from the same overall population. \*

## Appendix D

### Coding, Study, and Data Selection: Post-Eligibility Screening

The below paragraphs presented in bullet points describe relevant coding decisions not adequately covered by the coding manual.

#### Inclusion/Exclusion criteria post-initial screening

1. Studies that compare their own primary data with data from historical cohorts/archival datasets, and studies that compare cohorts from different time periods were excluded from both the systematic review and the quantitative synthesis (e.g., example study). This decision was made because such comparisons are likely to confound findings due to the influence of, for example, historical factors on the effect size data.

#### Study and sample characteristics

1. For some studies the record type could not easily be identified as either unpublished manuscripts or as government reports (e.g., some studies by the Van Voorhis et al. research group). If these records were exclusively available from a university website, they were coded as “university reports”.
2. Some studies reported (sub)sample(s) demographics for the (sub)sample(s) involved in recidivism analyses ( $k =$  ), while other studies reported these characteristics only for the original overall (sub)sample(s) before attrition at follow-up ( $k =$  ) or after attrition at follow-up ( $k =$  ). In addition, for those studies that reported demographics for the overall (sub)sample(s) after attrition at follow-up, analyses sometimes involved (a) (sub)sample(s) that was/were further reduced (e.g., due to missing data;  $k =$  ). In other words, in terms of representing the offenders (samples) involved in recidivism analyses, the accuracy of the sample characteristics varied widely between studies. During the coding process, it was not clear which sample characteristics would be reported by the largest number of studies, or which set of characteristics would be most relevant to report. Thus, when more than one set of sample characteristics were available from a single study, all were initially coded. In cases wherein there were separate effect sizes for subsamples, sample characteristics for all subsamples were coded when available, and the different sets of characteristics pertaining to each subsample were entered into the SPSS dataset.
3. Some studies reported effect size data pertaining to a number of separate mental disorders, without reporting the equivalent data for any/all mental disorders. For these studies, I did not calculate percentages with any mental disorder or proportions of offenders with any mental disorder who did or did not recidivate, because in all cases, a number of relatively common mental disorders were often excluded from the studies, making proportions with “any disorder” uninformative (e.g., if authors reported effect sizes and proportions of the sample(s) with for ADHD, PTSD, and Schizophrenia by recidivism status, combining the proportions with any mental disorder would not result in

effect sizes that truly represent the proportions of recidivists and non-recidivists who were mentally disordered, and thus, they were not combined into an overall “any disorder” subsample). Nevertheless, for studies that actually reported effect size data for “any mental disorder”, these effects were initially coded, along with which disorders were included in the authors’ definition of any mental disorder for the purposes of reporting in the qualitative review. The same strategy was employed for studies that reported effect size data for “any severe mental disorder”.

4. Almost all studies reported sample characteristics for a reduced total and/or sub-sample. For example, the total sample after attrition at follow-up may be  $N=300$ , while the sample size with data on the “mean age” variable may be  $n = 297$ . For almost all studies, the number of cases that had data on sample characteristics variables was reduced relative to the overall sample, and tended to vary between specific sample characteristics as well (e.g., a study may have  $N=300$  participants after attrition at follow-up, 297 offenders with data on mean age, and 200 offenders with data on average risk level). This variation in sample size(s) was not accounted for in the coding scheme or in analyses.
5. I did not create a variable that accounted for any overlap (or exclusivity) of various diagnoses (i.e., whether participants could have more than one mental disorder), because only one study used other mental disorder diagnoses as (sub)sample exclusion criteria.
6. In the original coding sheets, only the percentages of the (sub)sample(s) with specific (single) mental disorders were coded, unless percentages with broader categories of mental disorders such as mood disorders were also explicitly reported (in which case both specific disorders and broad categories were reported). However, upon entry into the dataset, I also attempted, wherever possible, to combine percentages with various specific mental disorders into broader categories. However, in many cases, authors reported only one or a small, non-inclusive subset of the disorders belonging to a single broad category (e.g., for mood disorders, only the percentage with major depression was reported, to the exclusion of bipolar disorder). In such cases, the broader category variable was left blank. However, if all the main/most common disorders included in a broader disorder category were reported, these were combined and reported under the appropriate broader disorder category. The following heuristics were developed to allow for this: In order to code the percentage of the sample with “any mood disorder”, authors had to report the percentage of offenders with major depression and bipolar disorder. To code “any anxiety disorder” authors had to report the percentage of offenders with GAD, OCD, and PTSD. To code any psychotic disorder, authors had to report percentage for at least all schizophrenia spectrum disorders. Of note, almost no studies reported the percentages necessary to code the broadest possible categories of mental disorders: internalizing and externalizing disorders. Thus, these broad categories were not coded in the dataset and were not used in analyses.
7. When the (sub)sample(s) consisted of all mentally disordered offenders (legally labeled MDOs housed in forensic mental health hospitals or secure units), the percentage receiving mental health treatment was always coded as 100%, given that all these offenders were very likely to have received some kind of mental health treatment while incarcerated.

### **Risk of bias variables**

1. The risk of bias variables pertaining to the representativeness of the subsamples were coded with the relevant (sub)sample type in mind. Specifically, if the relevant sample consisted entirely of offenders in a pre-trial jail, coders evaluated whether the sample was representative of the greater population of pre-trial detainees rather than the entire women offender population in general.

### **Effect size data**

1. Effect sizes and related data derived from multivariate analyses were coded, despite being excluded from quantitative syntheses. This was done in order to ease reporting of multivariate findings in the qualitative review (i.e., including this data allowed for the calculation of averages, variances (i.e., standard deviations), and frequencies of, for example, study and sample characteristics for multivariate analyses).
2. If analyses reported in the study only controlled for “time at risk”, and otherwise included no additional variables besides the relevant mental health predictor, the analyses were coded as univariate and considered eligible for inclusion in the quantitative synthesis of univariate effects.
3. A mental health predictor with more than two response categories, such as categorical mental health variables with three or more response options, were coded as continuous.
4. For both effect size and other variables, all numbers were reported with three decimals whenever possible. Decimals were always rounded up. For example, 4.4655 would be coded 4.466.
5. Means and standard deviations on continuous mental health variables were not reported in any included study. In one study, only means were reported without any index of variance. Thus, the initially included means and standard deviations variables were removed from the SPSS dataset.

### **Study-specific coding and data entry decisions**

1. Blanchette (1996) and Blanchette and Motiuk (1996)
  - The CSC originally conducted an initial study on the exact same sample as the one used in these two (completely overlapping) studies. The only difference in samples is that CSC included one additional case (i.e.,  $N = 77$  vs.  $N = 76$ ). The CSC study was internal and was not available for this meta-analysis. In Blanchette (1996), information regarding the mental health of the women in the sample originating from CSC’s original study was presented in the introduction. Although this information was included in the current author’s original coding sheets, it was not entered in the SPSS dataset because of the difference of one case in the sample sizes. Meaning, there were no effect size data available for this sample.
  - In the SPSS dataset, I did not fill in the mental health disorder categories variables for this study. I chose not to enter this because there was no fair measure of the percentage of offenders with a diagnosis falling into the diagnostic categories (see

bullet point five in the sample characteristics paragraph of this appendix). For example, although the studieffect size measured proportion of women with major depression, other key/important diagnoses in the mood disorder category were not measured (e.g., bipolar disorder). Similarly, the studieffect size measured GAD only, among the many anxiety disorders. As apparent from the data from the original CSC study that was presented in Blanchette (1996), including only one diagnosis from a disorder category/type would not at all represent the true proportions of women who suffered from diagnoses falling into specific diagnostic categories.

## 2. Scott et al. (2014; 2016)

- These studies (particularly 2014) contained a huge number of effect sizes. All were coded. There is one particularly important question that needs to be answered before these effect sizes can be included in analyses. The study authors report that they examined, at least for some effect sizes, time to recidivism (i.e., their recidivism measure was continuous). This initially caused confusion, because they report only Odds Ratios (ORs) and 95% CIs for all analyses. It seems possible that recidivism was measured dichotomously for the first effect sizes with a follow-up time of 36 months, while the remaining effect sizes that pertain to certain time periods (e.g., recidivism between the 1<sup>st</sup> and 3<sup>rd</sup> month of follow-up) and the effect sizes from Scott (2016) represent time to recidivism analyses. Upon further careful examination of the studies, it became clear that the authors did intend to report ORs, despite the continuous nature of the recidivism outcome variable. The current author researched the matter, and found support for the contention that hazard ratios (HRs) can in fact be converted into ORs. It was therefore concluded that the reporting of ORs for time to recidivism analyses is acceptable. The effect sizes were coded as ORs.
- It appeared that “traumatic stress” was likely to be equivalent to a PTSD diagnosis, and thus was coded as PTSD.
- Scott et al (2014) report recidivism for 36 months follow-up for their entire sample followed by the recidivism rates at one to three months, four to 12 months, and so on. I entered effect sizes for all follow-up times. However, the researchers also reported sample characteristics for recidivists at one to three months, four to 12 months, and so on, as well as characteristics for the non-recidivists. The samples of recidivists and non-recidivists make up the effect size (ES) for the relevant follow-up time. It does not make much sense to enter the sample characteristics for the recidivists and non-recidivist groups separately, as no other study reports the subsample characteristics for recidivists and non-recidivists. Nevertheless, having data for each sample involved in each analysis (each ES) for each follow-up time, would be beneficial. Thus, I combined the sample characteristics of the recidivists involved in the particular ES (e.g., those who recidivated between one and three months) with the characteristics of the non-recidivists involved in the same ES. In conclusion, the sample characteristics for each ES reported in this study were coded.
- All mental health variables were continuous, so they were coded as continuously measured mental health variables, and as MH Indicators/predictors rather than diagnoses.

3. Mannerfelt and Håkansson (2018)

- The study authors reported the percentage of the sample presenting with “cognitive problems”. This percentage was not coded as cognitive disorders in the dataset, because it was not clear that the problems necessarily warranted a diagnosis of a cognitive disorder. Instead, it was coded as another mental health predictor variable and not a diagnosis. The same data entry strategy was employed for “hallucinations” in this study.

4. Ducat et al. (2017)

- Intellectual disability was coded as a cognitive disorder.
- The majority, but not the full sample, were in the community at the time of mental health assessment, so the setting in this study was coded as “mixed correctional facilities and community”.
- Childhood behavioural disorder classified and coded as “any childhood disorder”.

5. King et al. (2018)

- Authors reported both the percentage with major depression and the percentage with bipolar disorder. Thus, the proportion with any mood disorder was coded by adding the percentages with depressive disorder and bipolar disorder.
- The authors reported the percentage of offenders with various types of what the authors referred to as “severe mental disorder”. However, it was not clear whether the percentages pertained to the total sample, or the subsample of offenders with a severe mental disorder. Neither possibility explains the proportions reported in Table 2 in the original study. Careful inspection by the current author as well as the secondary coder resulted in conclusion that the percentages reported cannot pertain to either the overall mixed sample or to the subsample with severe mental disorder. Thus, no proportions were coded for this study.

6. Tripodi et al. (2019)

- This study used the criterion “currently experiencing at least one psychotic symptom” for diagnosis of a psychotic disorder. This was similar to another study, in which the percentages with “hallucinations” were provided. In both cases, I coded these predictors as other mental health indicators (predictors) rather than as proper diagnoses, because according to DSM-IV-TR criteria, a diagnosis of a psychotic disorder cannot be made based solely on the presence of hallucinations. Given that at other studies also reported psychotic symptoms, a response option for the other mental health indicator/predictor variable in the dataset called “psychotic symptoms” was created, in which the predictor for the current study was coded.

- The authors reported the percentage of offenders who identified as “Pacific Islanders” or “Asian” as a single ethnicity. Thus, in the dataset, Pacific Islander was coded under the percentage of Asian offenders.

8. Harris et al. (2002)

- Clinicians’ ratings on a checklist were used to measure self-harm. I coded this as a clinical interview in the assessment method variable.
- There was the question of whether, for samples of mentally disordered offenders (MDOs), I should code the proportion with any mental health disorder as 100%. I decided that I should, because in the few studies of MDOs that were coded, the offenders without an axis 1 disorder were psychopathic. Given that psychopathy and antisocial personality disorder (APD) are considered closely related and/or greatly or entirely overlapping, this would result in the most consistent approach.

9. Van Voorhis et al. (2007; 2008; 2009; 2010)

- In some cases, there were two sets of risk levels (with percentages) reported for the same sample. One set was the original risk levels, while the other was the adjusted risk levels estimated based on the new instrument the authors had created. In such cases I coded both sets of risk levels initially, but entered only the original risk levels in the dataset.
- Sometimes there were two sets of percentages of offenders with violent and non-violent index offenses. I coded both sets but only entered the set that pertains to the violent and non-violent offenses as normally defined (i.e., ones most compatible with the definitions of violent vs. non-violent offenses in other studies).
- In terms of ethnicities, “Pacific Islanders” were included in the percentage of Asian ethnicity.
- In some cases, one document reported effect size data, while another provided additional study or sample characteristics, or risk of bias, and in other cases the same effect sizes for the same samples were reported in two or more studies. If several documents reported the same effect size, the “YEARMOD” variable in the dataset (i.e., the variable created for potential moderation analyses) was set to the earliest study.
- One mental health predictor was called “psychosis or suicidal”, I coded this using a response option tailored to this predictor and as a mental health indicator/predictor rather than as a diagnosis.
- Not all recidivism outcomes had a base rate. Thus, some effect sizes were entered into the dataset without a base rate.
- For Van Voorhis et al. Missouri Prison Sample ( $N = 272$ ) differently structured risk levels were reported alongside standard risk levels. I coded both but entered only the standard risk levels.
- Van Voorhis et al. Missouri Probation Sample ( $N = 313$ ) reported different percentages of ethnicities for different (sub)sample sizes. In principle, I would enter the ethnicities associated with the largest most complete sample size, but in this case

the participants' ethnic background were so much more detailed for the subsample of 265 offenders that I reported these instead, as they would be more informative.

- For the Maui Probation Sample I left the variable "percentage non-Black minority" (this variable was only included in the dataset, not in the coding sheets) blank, because all ethnicities, including whites, were in a minority in this sample.

#### 10. Chang et al. (2015)

- I coded and entered all possible effect size (ES) data (with the exception of effect sizes from model 3, which were deemed to be too esoteric for the current review). from Chang et al. (2015). In one instances, the authors reported probabilities for each subgroup separately (i.e., separately for the mental health (MH) and non-mental health (NMH) subgroups the authors were comparing in the time-to-recidivism analyses). In another instance, the authors reported median months to recidivism, again separately for the MH and NMH subgroups. In order to accommodate these data, I had to create two new possible responses for the type of coefficient/ES variable, as well as add another eight new variables that would cover the probability/median months "coefficients" of the MH and NMH groups, and the upper and lower confidence intervals for both of these subgroups for the two analyses in question. However, provided the availability of a large number of standard effect sizes, these coefficients were not included in analyses.
- Chang et al. presented a category of disorders that they named "other (non-ADHD) developmental or childhood disorder". I entered the associated sample proportion for this category in the existing variable "Proportion of Non-ADHD Childhood, Infancy, and Adolescence Disorders", as it seemed unlikely that at least two other studies would use the label utilized by Chang et al in a way that would allow aggregation. I therefore put the proportion under the existing variable that best covered the relevant cluster of disorders.
- It was reported that anxiety disorders were the most common types of disorders. However, the authors did not report percentages suffering the specific varied diagnoses that belong under the label "anxiety disorders". Therefore, in the dataset, the proportion (15%) was reported only under "most common diagnostic category", while the "most frequent specific diagnosis" variable was left blank.
- The authors used the term "schizophrenia spectrum disorders" but used the term incorrectly. The schizophrenia spectrum ordinarily refers to the previously separate schizophrenia types (e.g., catatonic vs. paranoid). Had this been the defined of schizophrenia spectrum disorders in this study, I would have entered the proportion of offenders with this disorder (i.e., one of its variants) under "Proportion Schizophrenia". However, I believed the authors did not define schizophrenia spectrum disorders in the correct way. Instead, schizophrenia spectrum disorders also included delusional disorder and schizoaffective disorder. For this reason, I opted to enter the proportions/entries for this disorder category as "psychotic disorder" throughout the dataset and left the "Proportion Schizophrenia" variable blank.

#### 11. Davies et al. (2007)

- In this study, the mental health group was the women with a non-personality (axis I) mental disorder while the non-mental health group consisted of women without an axis I disorder but with psychopathic personality. Given the limitations of the conclusions that could be made regarding the role of mental disorder (axis I) in the maintenance of criminal behaviour based on this comparison, this study was included in the qualitative review but excluded from quantitative synthesis.

12. Light et al. (2013)

- Authors reported a large number of ethnicities (e.g., four “white” ethnicities). In data entry, I combined all the categories of white and calculated an overall percentage that were white, and did the same for the mixed background, Asian background, Black, and Other categories. I also amended the coding sheets to demonstrate how these categories were combined.
- It was not clear from the study whether the setting was jail or prison. Sentences varied from one month up to four to five years however, so it was natural to assume that there were participants recruited from different settings (e.g., in Canada, provincial jails house, among others, offenders with sentences shorter than two years less one day, while federal prisons house offenders with sentences two years less one day or longer). Given that all offenders were incarcerated, I entered the setting as “unspecified correctional facility (jail and/or prison)” in the dataset.
- I did not enter the percentage of women who had been treated for a mental health problem in the 12 months prior to custody because the authors did not specify whether it was mental health treatment or substance abuse treatment, or both.

13. Steels et al. (1998)

- The proportion of offenders with any disorder was set to 100% in order to keep consistent with the other studies of mentally disordered offenders (MDOs). All participants had either a psychotic disorder, were neurotic, or were psychopathic.
- Authors reported the proportion of the sample that was “neurotic”. To accommodate this term, I added “neurosis” as a mental health condition/disorder along with the rest of the collection of possible mental health diagnoses and categories in the dataset.
- In order to accommodate legal classification as a mental health assessment method, I added it as a response option to the mental health assessment method variable in the dataset. This was done to accommodate the MDO studies. This response option was only used if there was no additional information regarding the protocol/assessment tools originally used to derive the diagnoses for MDOs.

14. Erickson (2014; 2016)

- The relevant effect sizes reported in Erickson (2016) could not be determined to have resulted from multivariate or univariate analysis. In fact, the analyses presented were, to be best of the current author and the secondary coder’s knowledge, uninterpretable. Given that the original author could not be reached,

the effect size data, but not the contextual data points were excluded from analyses, which involved only effect sizes from the 2014 study (i.e., only sample characteristics were gleaned from Erickson, 2016).

- For the main overall sample ( $N = 300$ ; 292-294 in analyses), I coded the sample characteristics as though they applied “before attrition” in the dataset, because some of the characteristics reported (e.g., age) were applicable to the original  $N = 300$  women. However, many of the sample characteristics applied to the reduced sample of 294 to 292 offenders. For the subsamples GO\_DOC and GO\_HB from 2016 document, I coded the available sub-sample characteristics pertaining to each of these groups and specified that the sample characteristics applied to those subsamples after attrition, as all the available characteristics reported for this sample pertained to the samples involved in analyses, rather than the original samples.
- I added a new category of “setting” specifically for this study, as it differed for this study relative to all the other studies coded. The offenders were housed in a private correctional assessment and treatment facility in New Jersey.
- The study author reported the percentage of the sample that was prescribed psychotropic medications and this data was initially coded with a cautionary note. The problem is that there is no way to determine whether the offenders actually adhered to their medication regimes or not, and for this reason, I did not enter this information in the dataset.
- The study examined the following mental health variables; BPD, anxiety, and depression, as well as suicidal ideation. The disorders were identified in the correctional assessment and rehabilitation center, but rather than interviews/DSM/ICD-based diagnoses, they were made based on self-report on the Personality Assessment Inventory (PAI). The scores on the PAI were continuous. Nevertheless, I coded and entered these variables as continuous “diagnoses” and “diagnostic categories” rather than as “other mental health indicators/predictors” in the dataset.

#### 15. Van Voorhis et al. (2012)

- For this study, the measures of mental health conditions were continuous and based on correctional risk/need assessments. For some effect sizes (ES), the depression scale was collapsed into three categories (low, medium, high), while the original depression scale was kept intact for other ES. Given that in this study, the authors were basically measuring symptoms of certain mental health conditions (e.g., psychotic symptoms were assessed with two items), I considered these mental health predictors (e.g., depression collapsed, depression symptoms, psychotic symptoms) as other mental health indicators/predictors, rather than strict diagnoses/diagnostic categories.
- The sample characteristics reported for the overall all general offenders (GO\_ALL;  $N = 626$ ) sample were only available for a subset of offenders in this sample ( $n = 403$ -556), and the attrition rate was extremely low ( $n = 1$  lost at follow-up), I coded the “accuracy of demographics” variable as though the characteristics were reported for the total sample after attrition at follow-up, even

though it cannot be ascertained that the one offender who was lost at follow-up was still included in the sample characteristics reported for this overall sample.

- Page 41, Table 11 of the original study has the superscript “a” for correlations that excluded the Missouri cases. In the table notes, it is indicated that this superscript means that when reconvictions at six months was the outcome, the Missouri sample ( $N = 313$ ) was excluded, leaving only 312 cases. It does not mention anything about the Missouri sample being excluded from the re-arrests at six months as the outcome. Despite this, there is an “a” superscript on the header “arrests” in the second column of the table. I chose to assume that this superscript had been attached to “re-arrests” erroneously and that the sample for re-arrests at six months was still 625. This seems more likely than that the authors forgot to mention the exclusion of the Missouri cases in the table note.
- To accommodate the unique recidivism outcomes for this study, I added new response option for “Recidivism definition” variable (i.e., “offense fail” = re-arrests, re-convictions, or registered violations [whether violation was legally pursued or not]; “any fail”).
- Three of the effect sizes originally coded for this study were excluded from the dataset (i.e., they were not entered), because the subsample sizes and the number of offenders who recidivated were not reported for these subsamples, making it impossible to aggregate these effect sizes with those of the remaining studies.

**Appendix E**

**Meta-analytic Formulae**

**Table 1E.**

*Formulae that were used in the Current Systematic Review and Meta-analysis*

Value or Metric to be Calculated	Data available
<b>SE of log OR</b>	$V_{logOR}$ (variance of logit/log OR)
$\sqrt{V_{OR}}$	
<b>Standard error (SE) of the log OR ( )</b>	2x2 table/proportion/frequencies
$SE\{\ln(OR)\} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$	
<b>Confidence intervals (CI) for OR</b>	SE of log OR (when 2x2 table is unknown)
Lower 95% CI = $e^{(\log OR - 1.96 * SE_{log OR})}$ Upper 95% CI = $e^{(\log OR + 1.96 * SE_{log OR})}$	
<b>Confidence intervals (CI) for logit (log OR)</b>	logit and SE of log OR (2x2 table is unknown)
Lower 95% CI = $\log OR - 1.96 * SE_{log OR}$ Upper 95% CI = $\log OR + 1.96 * SE_{log OR}$	
<b>Standard error (SE) of log OR</b>	Confidence intervals of OR
$SE_{log OR} = \frac{LN(OR_{upperlimit}) - LN(OR_{lowerlimit})}{3.92}$	
<b>Cohen's d</b>	Logit (log OR)
$d = \text{Log}(OR) * \left(\frac{\sqrt{3}}{\pi}\right)$	
<b>Cohen's d from t-test</b>	(Only) t-value and total N
$d \approx \frac{2 * t}{\sqrt{N}}$	
<b>Logit (log OR)</b>	Cohen's d
$OR = d * \frac{\pi}{\sqrt{3}}$	

**Appendix E**

**Meta-analytic Formulae**

**Table 1E.**

*Formulae that were used in the Current Systematic Review and Meta-analysis (Cont.).*

Value or Metric to be Calculated	Data available
<b>Variance of Cohen's <i>d</i></b>	Logit (log <i>OR</i> )
$d_{logOR} = v_{logOR} * \left(\frac{3}{\pi^2}\right)$	
<b>Confidence intervals (CI) of <i>d</i></b>	Standard deviation of <i>d</i> ( <i>SD<sub>d</sub></i> )
$95\% \text{ CIs} = d \pm (1.96 * SD_d)$	
<b>Standard deviation (<i>SD</i>) of <i>d</i></b>	Variance of <i>d</i> ( <i>d<sub>var</sub></i> )
$\sqrt{V_d}$	

*Note.* The formulae in this table were derived from the following sources: Hedges, L. V., & Tipton, E. (2010). Chapter 58 Meta-Analysis. In A. Steptoe (ed.), *Handbook of Behavioral Medicine* (pp. 901-921). Springer Science+Business Media. <https://doi.org/10.1007/978-0-387-09488-8>; Helmus, L., & Babchishin, K. (May, 2012b). *How to guide: Writing your coding manual*. Ottawa, ON: Carleton University; Polanin, J. R., Snilstveit, B. (2016). *Campbell Methods Policy Note on Converting Between Effect Sizes* (Version 1.1, updated December 2016). Oslo: The Campbell Collaboration. <https://doi.org/10.4073/cmpn.2016.3>; Researchgate online forum - August 6th, 2020 [https://www.researchgate.net/post/Can\\_anyone\\_suggest\\_a\\_formula\\_for\\_converting\\_variance\\_of\\_log\\_odds\\_ratio\\_into\\_standard\\_error\\_of\\_log\\_odds\\_ratio\\_in\\_meta-analysis](https://www.researchgate.net/post/Can_anyone_suggest_a_formula_for_converting_variance_of_log_odds_ratio_into_standard_error_of_log_odds_ratio_in_meta-analysis); Hintze (2008). *Chapter 218 Confidence Intervals for the Odds Ratio of Two Proportions*. PASS User's Guide III. Kaysville, Utah: NCSS; Helmus, L., & Hanson, K. R. (2011). More fun with statistics! How to use logistic regression to predict criminal recidivism risk. *Crime Scene, 18*(2), 7-12; Higgins, J. P. T., Li, T., & Deeks, J. J. (2019). Chapter 6: Choosing effect measures and computing estimates of effect. In J.P.T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0. Cochrane. Retrieved August 6th, 2020, from <https://training.cochrane.org/handbook/current/chapter-06> (\*Section 6.3.1 of the handbook); Polanin, J. R., Snilstveit, B. (2016). *Campbell Methods Policy Note on Converting Between Effect Sizes* (Version 1.1, updated December 2016). Oslo: The Campbell Collaboration. <https://doi.org/10.4073/cmpn.2016.3>; Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1-12. <https://doi.org/10.3389/fpsyg.2013.00863>; Polanin, J. R., Snilstveit, B. (2016). *Campbell Methods Policy Note on Converting Between Effect Sizes* (Version 1.1,

updated December 2016). Oslo: The Campbell Collaboration. <https://doi.org/10.4073/cmpn.2016.3>; Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2-18. <https://doi.org/10.1037/a0024338>; Online Calculators: Lenhard, W., & Lenhard, A. (2016). *Calculation of Effect Sizes*. Retrieved July 29<sup>th</sup>, 2020 from [https://www.psychometrica.de/effect\\_size.html](https://www.psychometrica.de/effect_size.html). Dettelbach (Germany): Psychometrica. doi: 10.13140/RG.2.2.17823.92329; Wilson, D. B. (2001). Practical Meta-Analysis Effect Size Calculator [Online calculator]. Retrieved July 28<sup>th</sup>, 2020, from <https://campbellcollaboration.org/research-resources/effect-size-calculator.html>.; McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*(2), 361–365.

## Appendix F

### Individual Effect Size Selections: Documenting Judgment Calls

This document details the selection of individual effect sizes for utilization in main meta-analyses. A combination of syntax manipulation, re-codes of original variables in the dataset, and manual selection was employed. This method was preferred in terms of defensible time management. Given that manual selection was part of the process however, every selection decision is detailed in the document below.

#### *General Information*

- Prior to ES selection, a variable entitled “ES#” was created, adding a random number to all ES (all rows in the dataset). Decisions outlined below make use of this number system to allow for the identification of specific ES.
- Organization of document (excluding introductory paragraphs): 1) the overall mental health predictor with the total number of independent ES included in *italics*, any relevant overall commentary in *italic bullet form*, followed by study ES selections. In this latter section, the study citation is provided first, followed by the selected ES in **bold font**. When only the ES number (**ES#**) is provided, it means that only one ES was available thus automatically selected. Any additional information presented following the ES number details the criteria upon which the selection ultimately relied. This organizational pattern repeats for three categories of independent mental health predictors.
- The present documentation pertaining directly to selection decisions for ES is organized according to what predictor variables were used to identify individual mental health variables. I had created three main predictor variables consisting of large numbers of possible options to represent unique disorders and specific classes of predictors. First, selection decisions for the most specific, most exclusive overall variable are presented. Second, equivalent selection decisions for a more general, yet not all-inclusive variable are documented. Lastly, decisions pertaining to the most comprehensive, inclusive (i.e., least restrictive) variable are presented. In the most specific variable, Major depressive disorder may be considered separate from Dysthymia, the two not being combined into a single ES. In the general variable, the two may be combined into “any depressive disorder”, but may not be grouped into a broader “mood disorder” category. Finally, in the third variable, any and all indications of any mood related problems may be collected under the umbrella term “mood” and include binary, categorical, and continuous predictors, all mood disorders regardless of type, symptoms, and so on).

#### *Selection Rules and Guiding Principles*

- The ES was selected based on the following priorities: The ES associated with 1) the maximum sample size, 2) the maximum length of follow-up, and 3) the most comprehensive measure of recidivism should be the one selected when multiple options are available. There was a single exception, if the recidivism outcome with the highest predicted base rate consisted exclusively of technical violations or breaches associated with conditional release, a recidivism outcome with a lower expected base rate was preferred, all other factors held

constant. The reason for this decision was that the primary behaviour or outcome of interest are criminal offenses and less so behaviours criminalized on the sole grounds that the individual has previously been in contact with the justice system (e.g., the act of stepping outdoors after 10:00pm is only criminal if it constitutes a violation of conditional release).

- Importantly, the criteria listed here are not necessarily in order of importance. In some cases, the maximum sample size was associated with an arguably too modest follow-up period (e.g., six months) while a sample reduced in the most minor way provided a follow-up of three years. In such a case, all other factors held constant, the ES with the superior follow-up time and thus the most advantageous base rates was most likely to be selected.
  - *Of note*, there were cases in which applying all three selection criteria still left multiple potential effect sizes available for selection. In such cases, I attempted to view the context of the effect size in its entirety, and make an educated judgment call based on this careful inspection. Unfortunately, some subjectivity was entirely unavoidable (e.g., Van voorhis and colleagues', 2008 to 2010 studies produced hundreds of effects that were almost or in fact wholly indistinguishable, without a clear and objective way to determine the superior effect size). In such instances, I have provided detailed notes for each study below, making the process as transparent as possible. Most commonly, other contextual factors had to be considered as part of the selection process. For example, when no selection criteria clearly favored on ES over one or multiple others, factors such as the relative representativeness of the predictor associated with the ES could come into play. For example, in the broadest, most inclusive and comprehensive “*Mood (Any; symptoms, scores, diagnoses, or risk of)*” predictor, an ES associated with a general binary “mood disorder” predictor may be selected in favor of isolated ES that represented a predictor only used in a single study and that was expected to differ most sharply from the remaining effects. In short, a strategy aiming to reduce the risk of outlying values obscuring the actual effect or resulting in unacceptably large confidence intervals and thus a highly imprecise estimate, could be employed
- In selecting an ES for an individual predictor, the measurement or scaling of the predictor was not taken into account. An ES was selected without taking measurement into account, and only after, in conjunction with analyses, were effect sizes pertaining to binary vs. continuous predictors split to allow for analyses constrained to one type of measurement.

## Appendix F

### Individual Effect Size Selections Explained

#### Selection Set Number 1: The Most Specific Mental Health Index “PREDICTOR 1 SPEC”

##### Depression

*There were 11 independent ES.*

##### Selections by Study

Blanchette & Motiuk (1996): **ES# 6**: Most inclusive recidivism outcome (re-incarceration with technical violations).

Papadopolous (2011): **ES#24**: Maximum sample size and follow-up.

Scott et al. (2014; 2016): **ES#68**: Maximum sample size and follow-up.

Mannerfelt et al. (2018): **ES#98**

Tripodi et al. (2019): **ES#104**

Ducat et al. (2017): **ES#112**

Van Voorhis et al. (2007-2010): **ES#126**: Maximum follow-up time.

Light et al. (2013): **ES#168**

Erickson (2014; 2016): **ES#173**

Van Voorhis et al. (2012): **ES#218**: The largest sample size was 625, but the follow-up time for this maximum sample was as short as six months, making low base rates (not reported) highly probable. Given that an ES for a comprehensive recidivism measure (i.e., “*Any fail*”) was available for a sample only slightly reduced relative to maximum ( $n = 601$ ), with a more favorable 12-month follow-up, this ES was preferred.

Van Voorhis et al. (2013): **ES#244**: The largest sample size was 585, but the follow-up time was too short at six months, making low base rates (not reported) highly probable. Given that an ES for a comprehensive recidivism measure (i.e., “*Any fail*”) was available for a sample only modestly reduced from the total sample ( $n = 553$ ), with an improved 12-month follow-up period, the ES with a slightly reduced sample but with longer follow-up and better recidivism outcome was selected.

### **Alternative Depression Effects: Re-incarcerations as Outcome**

*There were six independent ES.*

In all but two cases, the desirable ES could be chosen based on sample size.

The ES selected were as follows: **ES#6**, **ES#24**, **ES#68**, and **ES#104**.

For Van Voorhis et al. (2012), the ES with the largest (marginally so) sample size had a problematically short follow-up of six months, so the ES with marginally reduced sample size ( $n = 187$  vs.  $N = 195$ ) with a longer follow-up of 12-months was selected (**ES#190**).

For Van Voorhis et al. (2013), the same issue arose, and the ES with minutely reduced sample size ( $n = 85$  vs.  $N = 91$ ) with 12-months follow-up was selected over the larger sample that was limited to six months follow-up (**ES#283**).

### **Alternative Depression Effects: 36 Months as the Follow-up of Choice**

*There were three independent ES.*

All ES selected based on sample size (**ES#24**, **ES#73**, **ES#104**).

### **Mood Disorder Diagnoses**

*There were eight independent ES.*

#### **Selections by Study**

Blanchette & Motiuk (1996): **ES#6**: Two ES for depression had maximum sample size and follow-up times. The reincarcerations including breaches recidivism outcome was selected as it was more inclusive than alternatives.

Papadopolous (2011): **ES#25**: This study posed a problem in terms of ES selection for mood disorders because there were ES for depressive disorders, bipolar disorder, and unspecified mood disorders. Given that the non-specific mood disorder variable seemed likely to include both bipolar and depressive disorders, it was selected as the most appropriate predictor. Two ES pertained to mood disorders not otherwise specified (NOS); the most inclusive option (reincarceration including technical violations) was selected.

Scott et al. (2014; 2016): **ES#68**: Maximum sample size and follow-up.

Mannerfelt & Håkansson (2018): **ES#98**

Tripodi et al. (2019): **ES#104**

Ducat et al. (2017): **ES#112**

Van Voorhis et al. (2007-2010): **ES#126**: Maximum sample size.

Erickson (2014; 2016): **ES#173**

- *Note that the ES selected were identical to those of the original inclusive depression predictor for all studies except Van Voorhis et al. (2012; 2013), which contained only correlations with symptoms or scores on mood disorder variables and Papadopolous (2011).*

### **Alternative Mood Disorder Diagnosis: Including Bipolar disorder Effects**

- *Initial attempts at creating an alternative “mood disorders” predictor simply lead to the selection of the same ES as the original depression variable with the exception of the ES for Papadopolous (2011), and the two ES for the Van Voorhis et al. (2012; 2013) studies, which were correlations. In response to this, I allowed for further experimentation and created an additional variable in which **ES#25** for Papadoplous (2011) was replaced with an ES with the most comprehensive recidivism outcome and improved base rate was ultimately selected (**ES#34**).*

### **Selection Set Number 2: The General Mental Health Indices “Predictor2GEN”**

#### **Mood disorder symptoms**

*There were only three independent ES available for this variable.*

- *This variable excludes binary diagnoses and focuses instead on just symptoms of (any) mood disorder.*
- *This variable included one ES from a study-specific, highly unusual and non-representative predictor entitled “risk of depression”. It was nevertheless deemed eligible for inclusion in order to maximize independent effects included.*

#### **Selections by Study**

Light et al. (2013): **ES#168**.

Van Voorhis et al. (2012): **ES# 220**: ES selected based on weighting of sample size vs. length of follow-up. The maximum sample size ( $N = 625$ ) was restricted to six months follow-up, while a very modestly reduced sample ( $N = 601$ ) being associated with a favorable 12-month follow-up period. Thus, length of follow-up was prioritized over minimal sample size gain. “*Any fail*” was the outcome.

Van Voorhis et al. (2013): **ES#244**: ES selected based on weighting of sample size vs. length of follow-up. The maximum sample size ( $N = 585$ ) was restricted to six months follow-up, while a very modestly reduced sample ( $N = 553$ ) being associated with a favorable 12-month follow-up

period. Thus, length of follow-up was prioritized over minimal sample size gain. “*Any fail*” was the outcome.

### **Anxiety Symptoms**

*There were four independent ES for this predictor.*

- *This variable included two unusual predictors; “anxiety and fear symptoms” and one “at risk of anxiety” reported in two studies. Despite the non-general nature of these predictors, they were considered eligible for inclusion.*
- *This variable excludes binary diagnoses but includes symptoms of PTSD.*

### **Selections by Study**

Scott et al. (2014): **ES#53**: Largest sample size.

Light et al. (2013): **ES#167**

Van Voorhis et al. (2012): **ES#183**: Maximum sample size and most comprehensive recidivism definition (“*any fail*”).

Van Voorhis et al. (2013): **ES#242**: Maximum sample size and most comprehensive recidivism definition (“*any fail*”).

### **Psychotic Symptoms**

*There were six independent ES.*

- *Note that there were an insufficient number of ES pertaining exclusively to binary diagnoses of psychotic disorders. Thus, only symptoms were examined as a predictor.*

### **Selections by Study**

Mannerfelt et al. (2018): **ES#102**

Tripodi et al. (2019): **ES#105**

Van Voorhis et al. (2007-2010): **ES#155**

Light et al. (2013): **ES#170**

Van Voorhis et al. (2012): **ES#179**: ES selected based on weighting of sample size vs. length of follow-up. The maximum sample size ( $N = 625$ ) was restricted to six months follow-up, while a very modestly reduced sample ( $N = 601$ ) being associated with a favorable 12-month follow-up

period. Thus, length of follow-up was prioritized over minimal sample size gain. “*Offense fail*” was the outcome.

Van Voorhis et al. (2013): **ES#250**: The largest sample size was only available for technical violations. I selected the next largest size, which constituted an enormous drop in size ( $n = 91$ ). Among the multiple remaining ES, the one with the most inclusive recidivism outcome was selected (reincarcerations with technical breaches included).

### **Self-harm (SH)**

*There were four independent ES.*

- *This self-harm variable included self-harm, suicidal ideation, suicide attempts, and homicidal and/or suicidal thoughts. Note that the SH variable for the third predictor (“Predictor3CUMU”), which is documented below, is identical to this one. An SH selection variable will therefore not be created for the third predictor.*
- *One study specific, unique predictor was included in this selection (“homicidal and/or suicidal thought”).*

### **Selections by Study**

Olson et al. (2016): **ES#21**

Scott et al. (2014): **ES#51**. Largest sample size.

Mannerfelt & Haakansson (2018): **ES#99**

Erickson (2014): **ES#174**

### **Selection Set Number 2: The Most Inclusive, Cumulative predictor “Predictor3CUMU”**

This predictor was the most inclusive of the three new predictor variables (i.e., all possible indicators of any given mental health issue were included, regardless of proximity, uniformity (whether it assessed exclusively the construct most obviously representative of the MH problem), and measurement/scaling.

### **Mood (Any)**

*There were 10 independent ES.*

- *This mood variable included symptoms of mood disorder without diagnoses, scores on measures of mood problems, formal diagnoses of any mood disorder, or indicators of an identified risk of developing a mood or mood regulation problem.*
- *This variable included one “at risk of depression” ES that was study specific and a-typical.*

### **Selections by Study**

Blanchette & Motiuk (1996): **ES#6**: Two ES had maximum sample size and follow-up time. ES selected based on best recidivism outcome (reincarcerations including breaches).

Papadoplous (2011): **ES#25**: Two ES available with the same sample size, recidivism definition, base rates, and follow-up times. Thus, the original predictor variable had to be examined in order to determine which one may be more suitable. One ES pertained exclusively to a diagnosis of Dysthymia, while the other pertained to mood disorders in general. The latter was selected as it more clearly represented the total selection of effects included.

Scott et al. (2014): **ES#68**: Largest sample size.

Mannerfelt & Haakansson (2018): **ES#98**

Tripodi et al. (2019): **ES#104**

Ducat et al. (2017): **ES#112**

Van Voorhis et al. (2007-2010): **ES#126**: Longest follow-up.

Light et al. (2013): **ES#168**

Van Voorhis et al. (2012): **ES#220**: Although a larger sample size ( $N = 625$ ) was available the follow-up period for that ES was only six months, so I selected the ES with the next largest sample size ( $N = 601$ ) and the most inclusive recidivism outcome (“*Any fail*”) and 12 months follow-up instead.

Van Voorhis et al. (2013): **ES#244**: ES was selected based on it having the second largest  $N$  and 12 months follow up ( $N = 553$ ). This was preferred over  $N = 585$  with only six months follow-up.

### **Anxiety (Any)**

*There were 11 independent ES.*

- *This variable included two atypical, study-specific predictors: “at risk of anxiety” and “anxiety and fear symptoms”.*

### **Selections by Study**

Blanchette & Motiuk (1996): **ES#9**: Most comprehensive recidivism definition.

Papadoplous (2011): **ES#27**: Maximum sample size.

Scott et al. (2014): **ES#53**: Maximum sample size.

Mannerfelt & Haakansson (2018): **ES#100**

Tripodi et al. (2019): **ES#103**

Ducat et al. (2017): **ES#113**

Kubiak (2004): **ES#120**

Light et al. (2013): **ES#167**

Erickson (2014): **ES#172**

Van Voorhis et al. (2012): **ES#183**

Van Voorhis et al. (2013): **ES#242**. Largest sample size and most inclusive recidivism outcome (“*any fail*”).

### **Personality disorder (PD)**

*There were four independent ES.*

#### **Selections by Study**

Papadoplous (2011): **ES#23**

Scott et al. (2014): **ES#59**: There were four available ES, corresponding to Conduct Disorder Scale, Impulsive Personality Symptoms (SCID-II), Borderline Personality Symptoms (SCID-II), and Conduct Disorder Symptoms (SCID-II). There were no meaningful differences in sample size ( $N = 621 - 623$ ), base rates, recidivism definition, or follow-up times. Thus, I had to select the one most relevant to my research question. Given that my population includes only adult women, the relevance, and arguably the practice of assessing symptoms of conduct disorder, which is a childhood/adolescence-limited disorder, was questionable. Further, impulsive personality symptoms, seemed to me to lack specificity and may plausibly represent something other than a diagnosable personality disorder. In conclusion, borderline personality disorder (BPD) was selected as the ES for this study.

Ducat et al. (2017): **ES#115**

Erickson (2014): **ES#171**: Maximum sample size.

### **Psychosis**

*There were seven independent ES.*

#### **Selections by Study**

Mannerfelt & Haakansson (2018): **ES#102**

Tripodi et al. (2019): **ES#105**

Ducat et al. (2017): **ES#111**

Van Voorhis et al. (2007-2010): **ES#153**

Light et al. (2013): **ES#170**

Van Voorhis et al. (2012): **ES#223**: ES with slightly reduced sample size (i.e., difference of 22 participants), longer (12 months), and the most inclusive recidivism definition (“*offense fail*”) was selected (No other included recidivism option was available – only breaches).

Van Voorhis et al. (2013): **ES#260**: The maximum sample pertained only to technical violations at six months follow-up. A 12-month follow-up and “*any fail*” as recidivism outcome was preferred, despite the modest resulting sample of  $n = 85$ .

### **Psychiatric history**

*There were only three independent ES for psychiatric history and all were from different studies.*

- *This predictor consisted of ES for any variable that indicated any history of psychiatric care.*

### **Selections by Study**

Mannerfelt & Håkansson (2018): **ES#98**

Tripodi et al. (2019): **ES#104**

Ducat et al. (2017): **ES#109**

### **Any Mental Disorder (MD)**

*There were six independent ES.*

- *An additional MD variable was created for MD that was less restrictive so as to include major mental disorder (MMD, MD with and without comorbid MD, SA, or PD).*
- *One ES included in this variable was MD with comorbid SA.*

### **Selections by Study**

Blanchette & Motiuk (1996): **ES#1**. Most inclusive recidivism outcome.

Papadoplous (2011): **ES#26**: Largest sample size.

Visher & Bakken (2014): **ES#107**: Most inclusive recidivism outcome: “self-reported recidivism”.

Ducat et al. (2017): **ES#110**

King et al. (2018): **ES#118**

Van Voorhis et al. (2007-2010): **ES#123**: Longest follow-up time.

### **Anxiety disorder**

*There were six independent ES.*

#### **Selections by Study**

Blanchette & Motiuk (1996): **ES#9**: Most inclusive recidivism outcome.

Papadoplous (2011): **ES#42**: No differences in sample size, follow-up, or recidivism outcome. ES available specifically for OCD and for a more inclusive anxiety disorder predictor (i.e., diagnostic category). I chose the diagnostic category as it more clearly represented the population of effect sizes included in this predictor.

Mannerfelt & Haakansson (2018): **ES#100**

Tripodi et al. (2019): **ES#103**

Ducat et al. (2017): **ES#113**

Erickson (2014): **ES#172**: Maximum sample size.

### **PTSD**

*There were only four independent ES.*

#### **Selections by Study**

Papadoplous (2011): **ES#27**: Maximum sample size.

Kubiak (2004): **ES#120**

Van Voorhis et al. (2012): **ES#183**: Maximum sample size.

Van Voorhis et al. (2013): **ES#242**: Maximum sample size.

**Appendix G**

**List of Authors Contacted and Inquiry Results**

**Table 1G.**

*Authors Contacted for Information or Additional data pertaining to Included Studies or to Obtain Additional Studies*

Author	Contact info	Contact	About coded study?	# studies/data sets obtained	Univariate data obtained? (#)
<b>Bukten, Anne</b>	Researchgate	yes	no	0	No
Duggan, Conor; <b>(Steels, Mark)</b>	Researchgate	No	yes	0	No
<b>Janssen, Patricia A.</b>	patti.janssen@ubc.ca	yes	yes	0	No
<i>Lanes, Eric</i>	<i>could not be found</i>	<i>No</i>	<i>yes</i>	0	<i>No</i>
<b>Kubiak, Sheryl P</b>	spk@wayne.edu	No	yes	0	No
<b>Benda, Brent B.</b>	bbenda@ualr.edu	No	yes	0	No
<b>King, Erin A.</b>	eaking@fsu.edu	Yes.	yes	0	Yes (2)
<b>Becker, Marion A.</b>	becker@fmhi.usf.edu	Yes.	no	0	No
<b>Alarid, Leanne</b>	Leanne.Alarid@utsa.edu	No	no	0	No
<b>Cloyes, Kristin</b>	Kristin.cloyes@nurs.utah.edu	No	no	0	No
<b>Grann, Martin;</b> Danesh, John;	c.m.g@telia.com john.danesh@phpc.cam.ac.uk	Yes	no	0	No
<b>Fazel, Seena</b>	seena.fazel@psych.ox.ac.uk				
<b>Houser, Kimberly A.</b>	houser@rowan.edu	No	yes	0	No
<b>Jaffe, Adi</b>	adi@ucla.edu	Yes	no	0	No
<b>Maden, A.</b>	a.maden@ic.ac.uk	Yes.	no	0	No
<b>Mcmurran, Mary</b>	Researchgate	Yes	no	0	No
<b>Ostermann, Michael</b>	Michael.Ostermann@rutgers.edu	Yes	no	0	No
<b>Rezansoff, Stefanie N.</b>	sra20@sfu.ca	Yes	no	0	No
<b>Skeem, Jennifer L.</b>	skeem@uci.edu; Researchgate	No	no	0	No
<b>Torgersen, Lori J.</b>	Researchgate	No	no	0	No
<b>Wilson, Amy B.</b>	abw38@case.edu	No	no	0	No
<b>Lovell, David</b>	lovelld@u.washington.edu	No	no	0	No

**Appendix G**

**List of Authors Contacted and Inquiry Results**

**Table 1G.**

*Authors Contacted for Information or Additional data pertaining to Included Studies or to Obtain Additional Studies*

Author	Contact info	Contact?	About coded study?	# of studies/data sets obtained	univariate data obtained? (#)
<b>Harris, Victoria</b>	vharris@u.washington.edu	No	No	0	No
<b>Hartwell, Stephanie</b>	stephanie.hartwell@umb.edu	Yes	No	0	No
<b>Payne, Clive</b>	<i>not found</i>	<i>No</i>	<i>No</i>	0	<i>No</i>
<b>Gagliardi, Gregg J.</b>	gagliard@u.washington.edu	No	No	0	No
<b>Stadtland, Cornelis</b>	cstadtla@psy.med.unimuenchen.de	No	No	0	No
<b>Yoshikawa, Kazuo</b>	yoshikaw@ncnp.go.jp	No	No	0	No
<b>Farabee, David</b>	dfarabee@ucla.edu	Yes	No	0	No
<b>Lee, Donald Thomas</b>	leedont@adelphia.net	No	Yes	0	No
<b>Quinsey, Vernon L</b>	Researchgate	Yes	No	0	No
<b>Phillips, Helen K</b>	PhillipsHK1@cardiff.ac.uk	No	No	0	No
<b>Van Voorhis, Patricia</b>	patricia.vanvoorhis@uc.edu	Yes	No	0	No
Bauman, Ashley	Researchgate				
<b>Jamieson, Elisabeth</b>	elizabeth.jamieson@wlmht.nhs.uk	No	No	0	No
<b>Solomon, Phyllis</b>	solomonp@ssw.upenn.edu	Yes	No	0	No
<b>Monson, Candice</b>	Candice.M.Monson@Dartmouth.edu	No	No	0	No
<b>Schonbrun, Yael Chatav</b>	yael_chatav_schonbrun@brown.edu	No	Yes	0	No
Zhao, Min	drminzhao@gmail.com	No	Yes	0	No
<b>(Du, Jiang)</b>					
Meyer, Jamie P	jaimie.meyer@yale.edu	Yes	Yes	1	Yes (1)
<b>(Odio, C.D)</b>					
<b>Olson, David E</b>	dolson1@luc.edu	Yes	Yes	0	Yes (1)
<b>Fu, Jeannia</b>	jeanniafu@gmail.com	No	No	0	No
<b>Scott, Christy K</b>	cscott@chestnut.org	yes	yes	0	No

## Appendix H

### References for Rejected Studies and Reasons for Rejection

Baillargeon, J., Binswanger, I. A., Penn, J. V., Williams, B. A., & Murray, O. J. (2009).

Psychiatric disorders and repeat incarcerations: The revolving prison door. *American Journal of Psychiatry*, *166*(1), 103-109. doi: 10.1176/appi.ajp.2008.08030416

Reason for rejection: This was not a recidivism study

Bales, W. D., Nadel, M., Reed, C., & Blomberg, T. G. (2017). Recidivism and inmate mental illness. *International Journal of Criminology and Sociology*, *6*, 40-51. doi: 10.6000/1929-4409.2017.06.05

Reason for rejection: This study did not aggregate by gender.

Barnett, W., Richter, P., Sigmund, D., & Spitzer, M. (1997). Recidivism and concomitant criminality in pathological firesetters. *Journal of Forensic Science*, *42*(5), 879-883.

Reason for rejection: Did not aggregate findings by gender

Becker, M. A., Andel, R., Boaz, T., & Constantine, R. (2011). Gender differences and risk of arrest among offenders with serious mental illness. *The Journal of Behavioral Health Services & Research*, *38*(1), 16-28. doi: 10.1007/s11414-010-9217-8

Reason for rejection: This study did not contain codeable data. Only relevant analyses compared women with and without severe mental illness in terms of membership in various classes that differed in terms of recidivism rates and patterns. There were no data pertaining to women with no recidivism (i.e., all classes consisted of some recidivists).

Begun, A. L., Early, T. J., & Hodge, A. (2016). Mental health and substance abuse service engagement by men and women during community reentry. *Administration and Policy in*

*Mental Health and Mental Health Services, 43*, 207-218. doi: 10.1007/s10488-015-0632-

2

Reason for rejection: This was not a recidivism study

Benda, B. B. (2001). Factors that discriminate between recidivists, parole violators, and nonrecidivists in a 3-year follow-up of boot camp graduates. *International Journal of Offender Therapy and Comparative Criminology, 45*, 711-729. doi:

10.1177/0306624X01456006

Reason for rejection: All male sample

Benda, B. B. (2003). Survival analysis of criminal recidivism of boot camp graduates using elements from general and developmental explanatory models. *International Journal of Offender Therapy and Comparative Criminology, 47*, 89-101. doi:

10.1177/0306624X02239277

Reason for rejection: All male sample

Benda, B. B., Toombs, N. J., & Peacock, M. (2003a). An examination of competing theories in predicting recidivism of adult offenders five years after graduation from boot camp. *Journal of Offender Rehabilitation, 37*, 43-75. doi: 10.1300/J076v37n02\_03

Reason for rejection: All male sample

Benda, B. B., Toombs, N. J., & Peacock, M. (2003b). Discriminators of types of recidivism among boot camp graduates in a 5-year follow-up study. *Journal of Criminal Justice, 31*(6), 539-551. doi: 10.1016/j.jcrimjus.2003.08.004

Reason for rejection: All male sample

Bieber, S. L., Pasewark, R. A., Bosten, K., & Steadman, H. J. (1988). Predicting criminal recidivism of insanity acquittees. *International Journal of Law and Psychiatry*, *11*, 105-112.

Reason for rejection: Did not aggregate findings by gender; participants were not primarily offenders (only a small portion of the participants had committed a crime prior to receiving inpatient treatment).

Bjørkly, S., Sandli, C. S., Moger, T. A., & Stang, J. (2010). A follow-up interview of patients eight years after discharge from a maximum security forensic psychiatry unit in Norway. *International Journal of Forensic Mental Health*, *9*(4), 343-353. doi: 10.1080/14999013.2010.534698

Reason for rejection: Sample of women too small ( $N < 10$ ).

Bogenberger, R. P., Pasewark, R. A., Gudeman, H., & Bieber, S. L. (1987). Follow-up of insanity acquittees in Hawaii. *International Journal of Law and Psychiatry*, *10*, 283-295.

Reason for rejection: Sample of women was too small ( $n < 10$ )

Bonta, J., Pang, B., & Wallace-Capretta, S. (1995). Predictors of recidivism among incarcerated female offenders. *Prison Journal*, *75*(3), 227-294.

Reason for rejection: Did not contain any mental health predictors.

Brunton-Smith, I., & Hopkins, K. (2013). The factors associated with proven re-offending following release from prison: findings from Waves 1 to 3 of SPCR. Results from the Surveying prisoner crime reduction (SPCR) longitudinal cohort study of prisoners. England & Wales: Ministry of Justice Analytical Services. Retrieved from Ministry of Justice website <http://www.justice.gov.uk/publications/research-and-analysis/moj>

Reason for rejection: Did not contain any codeable data

Bukten, A., Lund, I. O., Rognli, E. B., Stavseth, M. R., Lobmaier, P., Skurtveit, S., Clausen, T., & Kunøe, N. (2015). The Norwegian Offender Mental Health and Addiction Study- Design implementation of a national survey and prospective cohort study. *Substance Abuse: Research and Treatment, 9*(S2), 59–66 doi: 10.4137/SaRt.S23546.

Reason for rejection: Did not contain any codeable data. The study authors were contacted and informed me that analyses had not yet commenced for this study.

Castillo, E. D., & Alarid, L. F. (2011). Factors associated with recidivism among offenders with mental illness. *International Journal of Offender Therapy and Comparative Criminology, 55*(1), 98-117. doi: 10.1177/0306624X09359502

Reason for rejection: Did not aggregate findings by gender.

Cloyes, K. G., Wong, B., Latimer, S., & Abarca, J. (2010a). Women, serious mental illness and recidivism: A gender-based analysis of recidivism risk for women with SMI released from prison. *Journal of Forensic Nursing, 6*(1), 3-14. doi: 10.1111 /j.1939-3938.2009.01060.x

Reason for rejection: Did not include a non-mentally ill or alternative adequate comparison group.

Cloyes, K. G., Wong, B., Latimer, S., & Abarca, J. (2010b). Time to prison return for offenders with serious mental illness released from prison. A survival analysis. *Criminal Justice and Behavior, 37*(2), 175-187. doi: 10.1177/0093854809354370

Reason for rejection: Did not aggregate findings by gender

Cobbina, J. E. (2010). Reintegration success and failure: Factors impacting reintegration among incarcerated and formerly incarcerated women. *Journal of Offender Rehabilitation, 49*(3), 210-232. doi: 10.1080/10509671003666602

Reason for rejection: This study was qualitative

Cook, M. R. (2014). *Mental health issues, self-concept, and interpersonal relationships as predictors of recidivism: Exploring correctional adaptation* (Doctoral dissertation).

Retrieved from ProQuest Dissertations & Theses Global. (UMI No. 3626522)

Reason for rejection: Included youthful offenders only

Daffern, M., Ferguson, M., Ogloff, J., Thomson, L., & Howells, K. (2007). Appropriate treatment targets or products of a demanding environment? The relationship between aggression in a forensic psychiatric hospital with aggressive behaviour preceding admission and violent recidivism. *Psychology, Crime, and Law*, 13(5), 431-441. doi: 10.1080/10683160601060861

Reason for rejection: Sample of women was too small ( $n < 10$ )

Danzer, A. (2003). *Erie County Community Reintegration of Offender with Mental Illness and Substance Abuse (CROMISA) program*. Erie, Pennsylvania, United States: Mercyhurst College Civic Institute

Reason for rejection: Did not aggregate findings by gender; correspondence with Mercyhurst College Civic Institute indicated that results were not aggregated by gender for any of their yearly reports

Dauphinot, L. L. (1996). The efficacy of community correctional supervision for offenders with severe mental illness (Unpublished doctoral dissertation). The University of Texas at Austin, Austin, Texas.

Reason for rejection: Did not aggregate findings by gender

Derkzen, D., Wardrop, K., & Wanamaker, K. (2019). *Women offender assessment: Can gender-informed variables improve risk prediction?* (Research Report R-413). Ottawa, Ontario: Correctional Service of Canada.

Reason for rejection: Did not aggregate findings by gender; study authors were not contacted because the odds of obtaining additional data from CSC seemed extremely low

de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health*, 3(2), 149-165. doi: 10.1080/14999013.2004.10471204

Reason for rejection: Did not aggregate findings by gender; sample of women too small

Du, J., Zhao, M., & Hser, Y-I. (2011). PO1-33-Drug-abusing offenders with co-morbid mental disorders: Gender differences in problem severity, treatment participation, and recidivism. *European Psychiatry*, 26, 33. doi: 10.1016/S0924-9338(11)71744-8

Reason for rejection: This is just a short, single page summary. It contains no data.

Edward, D. J., Farnham, F., Moorey, H., Lloyd, H., Hill, K., Blizard, R., & Barnes, T. R. E. (2002). *Outcome of psychiatric admission through the courts* (Occasional Paper No. 79). London, England: Great Britain Home Office Research, Development and Statistics Directorate, Communication Development Unit.

Reason for rejection: Did not aggregate findings by gender

Elbogen, E. B., & Johnson, S. C. (2009). The intricate link between violence and mental disorder: Results from the National Epidemiologic Survey on Alcohol and related conditions. *Archives of General Psychiatry*, 66(2), 152-161. doi: 10.1001/archgenpsychiatry.2008.537.

Reason for rejection: Did not use an offender sample

\*Erickson, J. R. (2016). *Relationships between abuse history, substance abuse, mental health and recidivism in female offenders* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (ProQuest No. 10252879)

Reason for rejection: The findings reported for this study were deemed to be unreliable.

Demographic/sample characteristics were retained from this study in order to provide these data for the effect sizes reported in Erickson (2014), which utilized the same sample.

Farabee, D., & Shen, H. (2004). Antipsychotic medication adherence, cocaine use, and recidivism among a parolee sample. *Behavioral Sciences and the Law, 22*, 467-476. doi: 10.1002/bsl.601

Reason for rejection: Did not aggregate findings by gender

Fazel, S., Wolf, A., Chang, Z., Larsson, H., Goodwin, G. M., & Lichtenstein, P. (2015). Depression and violence: a Swedish population study. *Lancet Psychiatry, 2*, 224-232. doi: 10.1016/S2215-0366(14)00128-X

Reason for rejection: The sample was a general population sample rather than offenders. No recidivism outcome measure.

Feder, L. (1991). A comparison of the community adjustment of mentally ill offenders with those from the general prison population: An 18-month follow-up. *Law and Human Behavior, 15*(5), 477-493.

Reason for rejection: All male sample

Ferguson, A. M., Ogloff, J. R. P., & Thompson, L. (2009). Predicting recidivism by mentally disordered offenders using the LSI-R: SV. *Criminal Justice and Behavior*, 36(1), 5-20. doi: 10.1177/0093854808326525

Reason for rejection: Did not aggregate findings by gender; no suitable mental health variables

Forry, J. B., Kirabira, J., Scholastic, A., & Rukundo, G. Z. (2019). Crime, recidivism and mental disorders among prisoners in Mbarara municipality, southwestern Uganda. *International Journal of Law and Psychiatry*, 62, 1-6. doi: 10.1016/j.ijlp.2018.10.006

Reason for rejection: Flaw in study design. Temporal order of mental health condition and recidivism event could not be established.

Freeman, N. J., & Sandler, J. C. (2008). Female and male sex offenders. A comparison of recidivism patterns and risk factors. *Journal of Interpersonal Violence*, 23(10), 1394-1413. doi: 10.1177/0886260508314304

Reason for rejection: Contained no mental health predictors

Fu, J. J., Herme, M., Wickersham, J.A., Zeleney, A., Althoff, A., Zaller, N. D., Bazazi, A. R.,...Altice, F.L. (2013). Understanding the revolving door: Individual and structural-level predictors of recidivism among individuals with HIV leaving jail. *AIDS Behavior*, 17(02), S145-155. doi:10.1007/s10461-013-0590-1

Reason for rejection: Did not aggregate findings by gender

Gagliardi, G. J., Lovell, D., Peterson, P .D., & Jemelka, R. (2004). Forecasting recidivism in mentally ill offenders released from prison. *Law and Human Behavior*, 28(2), 135-155. doi: 10.1023/b:lahu.0000022319.03637.45

Reason for rejection: Did not aggregate findings by gender

Gehring, K. S., & van Voorhis, P. (2014). Needs and pretrial failure: Additional risk factor for female and male pretrial defendants. *Criminal Justice and Behavior, 41*(8), 943-970. doi: 10.1177/0093854814538022

Reason for rejection: The mental health variable was unsuitable for the current study.

Specifically, the mental health predictor was a combination of a selection of mental disorders (e.g., ADHD) and mental health indicators (e.g., ever been hospitalized for a mental health problem)

Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the Historical part of the HCR-20. *Criminal Justice and Behavior, 27*, 97-114. doi: 10.1177/0093854800027001006

Reason for rejection: Contained no codeable mental health variables

Grann, M., Danesh, J., & Fazel, S. (2008). The association between psychiatric diagnosis and violent reoffending in adult offenders in the community. *BMC Psychiatry, 8*(92), 1-7. doi: 10.1186/1471-244X-8-92

Reason for rejection: Did not aggregate findings by gender

Grann, M., & Långström, N. (2007). Actuarial assessment of violence risk To weigh or not to weigh. *Criminal Justice and Behavior, 34*(1), 22-36. doi: 10.1177/0093854806290250

Reason for rejection: Did not aggregate findings by gender

Gray, N. S., Fitzgerald, S., Taylor, J., MacCulloch, M. J., & Snowden, R. J. (2007). Predicting future reconviction in offenders with intellectual disabilities: The predictive efficacy of VRAG, PCL-SV, and the HCR-20. *Psychological Assessment, 19*(4), 474-479. doi: 10.1037/1040-3590.19.4.474

Reason for rejection: No adequate mental health variables available

Gray, N. S., Snowden, R. J., Phillips, H., Taylor, J., & MacCulloch, M. J. (2004). Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: A comparative study of HCR-20, PCL: SV, and OGRS. *Journal of Consulting and Clinical Psychology, 72*(3), 523–530. doi:10.1037/0022-006X.72.3.523

Reason for rejection: There were no usable mental health variables

Gray, N. S., Taylor, J., & Snowden, R. J. (2008). Predicting violent reconvictions using the HCR-20. *The British Journal of Psychiatry, 192*, 384-387. doi: 10.1192/bjp.bp.107.044065

Reason for rejection: All male sample

Harmon, R. B. (2011). *Anti-stalking legislation, recidivism, and the mentally disordered stalker* (Doctoral dissertation). Retrieved from ProQuest Dissertation & Theses Global (UMI No. 3440384).

Reason for rejection: Did not aggregate findings by gender

Harris, V., & Koepsell, T. D. (1996). Criminal Recidivism in mentally ill offenders: A pilot study. *The Bulletin of the American Academy of Psychiatry and the Law, 24*(2), 177-186.

Reason for rejection: Did not aggregate findings by gender; sample of women < 10

Harris, V., & Koepsell, T. D. (1998). Rearrest among mentally ill offenders. *Journal of the American Academy of Psychiatry and the Law, 26*(3), 393-402.

Reason for rejection: Did not aggregate findings by gender

Hartwell, S. (2003). Short-term outcomes for offenders with mental illness released from incarceration. *International Journal of Offender Therapy and Comparative Criminology, 47*, 145-158. doi: 10.1177/0306624X03251093

Reason for rejection: Did not aggregate findings by gender

Hartwell, S. (2004). Comparison of offenders with mental illness only and offenders with dual diagnoses. *Psychiatric Services, 55*(2), 145-150. doi: 10.1176/appi.ps.55.2.145

Reason for rejection: Did not aggregate findings by gender

Her Majesty's Stationary Office (1975). *Report of the committee on mentally abnormal offenders*. London, England: Home Office, Department of Health and Social Security.

Reason for rejection: Did not contain any codeable data

Herrman, H., McCorry, P., Mills, J., & Singh, B. (1991). Hidden severe psychiatric morbidity in sentenced prisoners: An Australian study. *American Journal of Psychiatry, 148*(2), 236-239.

Reason for rejection: The study contained no recidivism data

Hickey, N., Yang, M., & Coid, J. (2009). The development of the Medium Security Recidivism Assessment Guide (MSRAG): An actuarial risk prediction instrument. *Journal of Forensic Psychiatry and Psychology, 20*(2), 202-224. doi: 10.1080/14789940802207683

Reason for rejection: Did not contain any mental health variables

Hiday, V. A., & Ray, B. (2010). Arrests two years after exiting a well-established mental health court. *Psychiatric Services, 61*(5), 463-468. doi: 10.1176/ps.2010.61.5.463

Reason for rejection: Did not contain any mental health variables

Hilving, R. (2010). *Violent female offending: examining the role of psychopathy and comorbidity with DSM-IV personality disorders* (Master's thesis). Retrieved from ProQuest Dissertations & Theses Global. (Accession No. 1487288)

Reason for rejection: Did not measure recidivism.

Hines-Randolph, N. L. (2014). *Without walls: An examination of women and the barriers impeding successful reintegration* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Accession No. 3629291)

Reason for rejection: All participants were recidivists (included no non-recidivists)

Ho, H., Thomson, L., & Darjee, R. (2009). Violence risk assessment: The use of the PCL-SV, HCR-20, and VRAG to predict violence in mentally disordered offenders discharged from a medium secure unit in Scotland. *Journal of Forensic Psychiatry and Psychology*, 20(4), 523-541. doi: 10.1080/14789940802638358

Reason for rejection: Did not assess any mental disorders other than psychopathy

Hodgins, S. (1983). A follow-up study of persons found incompetent to stand trial and/or not guilty by reason of insanity in Quebec. *International Journal of Law and Psychiatry*, 6, 399-411.

Reason for rejection: Did not aggregate by gender; participants were not offenders (only a small portion of the participants had committed a crime prior to receiving inpatient treatment).

Hodgins, S., Mednick, S.A., Brennan, P.A., Schulsinger, F., & Engberg, M. (1996). Mental disorder and crime. Evidence from a Danish cohort. *Archives of General Psychiatry*, 53, 489-496. doi: 10.1001/archpsyc.1996.01830060031004

Reason for rejection: This was not a recidivism study

Holtfreter, K., & Wattanaporn, K. A. (2013). The transition from prison to community initiative: An examination of gender responsiveness for female offender reentry. *Criminal Justice and Behavior*, 41, 41-57. doi: 10.1177/0093854813504406

Reason for rejection: This was a review article

Hopkins, K., & Brunton-Smith, I. (2014). *Prisoners' experience of prison and outcomes on release: Waves 2 and 3 of SPCR. Results from the Surveying Prisoner Crime Reduction (SPCR) longitudinal cohort study of prisoners*. England & Wales: Ministry of Justice Analytical Services.

Reason for rejection: Did not contain any mental health data/variables and did not aggregate findings by gender.

Houser, K. A., Saum, C. A., & Hiller, M. L. (2019). Mental health, substance abuse, co-occurring disorders, and 3-year recidivism of felony parolees. *Criminal Justice and Behavior*, 46(9), 1237-1254. doi: 10.1177/0093854819856924

Reason for rejection: Did not aggregate by gender

Håkansson, A., & Berglund, M. (2012). Risk factors for criminal recidivism-a prospective follow-up study in prisoners with substance abuse. *BMC Psychiatry*, 12. doi: 10.1186/1471-244X-12-111

Reason for rejection: Did not aggregate findings by gender

Håkansson, A., & Berglund, M. (2013). All-cause mortality in criminal justice clients with substance use problems-A prospective follow-up study. *Drug and Alcohol Dependence*, 132(3), 499-504. doi: 10.1016/j.drugalcdep.2013.03.014

Reason for rejection: No recidivism outcome variable

Jacoby, J. E., & Kozie-Peak, B. (1997). The benefits of social support for mentally ill offenders: Prison-to-community transitions. *Behavioral Sciences and the Law*, 15, 483-501.

Reason for rejection: Did not aggregate findings by gender; author not contacted because too large a proportion of the small sample was male.

Jaffe, A., Du, J., Huang, D., & Heser, Y-I. (2012). Drug abusing offenders with comorbid mental disorders: Problem severity, treatment participation, and recidivism. *Journal of Substance Abuse Treatment, 43*(2), 244-250. doi: 10.1016/j.jsat.2011.12.002

Reason for rejection: Did not aggregate findings by gender.

Jamieson, L., & Taylor, P. J. (2004). A re-conviction study of special (high security) hospital patients. *British Journal of Criminology, 44*, 783-802. doi: 10.1093/bjc/azh054

Reason for rejection: Did not aggregate findings by gender

Janssen, P. A., Korchinski, M., Desmarais, S. L., Albert, A. Y. K., Condello, L-L., Buchanan, M., ...Martin, R. E. (2017). Factors that support successful transition to the community among women leaving prison in British Columbia: a prospective cohort study using participatory action research. *CMAJ Open, 5*(3), E717-E723. doi:

10.9778/cmajo.20160165

Reasons for rejection: This study was rejected following direct communication with the lead author-Dr. Janssen. The reasons for rejection was threefold: (1) the order of the measurement of the predictor and the outcome was reversed, (2) the number of recidivists and non-recidivists could not be determined from the data presented (nor could it be provided by the lead author), and (3) poor communication of the nature of their results and apparently erroneous interpretation of their effect size for depression (the only mental health variable reported on; see detailed notes in the “meta-coding decisions evolving document” about this (potential) misinterpretation of the findings) left the coder (present author; Cathrine) doubtful about the quality and soundness of the researchers’ methodology, analyses, and interpretation and presentation of findings.

Karatzias, T., Power, K., Woolston, C., Apurva, P., Begley, A., Mirza, K., Conway, L., Quinn, C., Jowett, S., & Howard, R. (2017). *Criminal Behaviour and Mental Health*, 2, 72-84. doi: 10.1002/cbm.2043

Reason for rejection: No recidivism outcome measure

Kershaw, C., Dodswell, P., & Goodman, J. (1997). *Restricted patients: Reconvictions and recalls by the end of 1995: England and Wales*. Home Office Research, Development and Statistics Directorate.

Reason for rejection: Did not aggregate findings by gender; authors not contacted because study originated from the government

Komer, B., & Galbraith, D. (1992). Recidivism among individuals detained under a warrant of the Lieutenant-Governor living in the community. *Canadian Journal of Psychiatry*, 37, 694-698.

Reason for rejection: Sample of women  $n < 10$ .

Komer, B., Math, B., & Galbraith, D. (1992). Recidivism among individuals detained under a warrant of the Lieutenant-Governor living in the community. *Canadian Journal of Psychiatry*, 37, 694-698.

Reason for rejection: Sample of women was too small ( $n < 10$ )

Kröner, C., Stadtland, C., Eidt, M., & Nedopil, N. (2007). The validity of the Violence Risk Appraisal Guide (VRAG) in predicting criminal recidivism. *Criminal Behaviour and Mental Health*, 17, 89-100. doi: 10.1002/cbm.644

Reason for rejection: No suitable mental health variables; did not aggregate findings by gender

Langstrom, G. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior*, 34(22), 22-36. doi: 10.1177/0093854806290250

Reason for rejection: There were no appropriate (codeable) mental health variables

Lee, D. T. (2003). Community-treated and discharged forensic patients: an 11-year follow-up.

*International Journal of Law and Psychiatry*, 26, 289-300. doi:10.1016/S0160-2527(03)00039-6

Reason for rejection: Did not aggregate findings by gender

Lindsay, W. R., Elliot, S. F., & Astell, A. (2004). Predictors of sexual offence recidivism in

offenders with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 17, 299-305. doi: 10.1111/j.1468-3148.2004.00217.x

Reason for rejection: Did not contain any mental health variables

Lindsay, W. R., Steele, L., Smith, A. H. W., Quinn, K., & Allan, R. (2006). A community

forensic intellectual disability service: Twelve year follow-up of referrals, analysis of referral patterns and assessment of harm reduction. *Legal and Criminological Psychology*, 11, 113-130. doi:10.1348/135532505X55669

Reason for rejection: Did not contain any mental health variables

Lippman, K. R. (2006). *Retrospective analyses of relationships between nature of crime, mental*

*illness, history of mental health treatment, and recidivism in a San Francisco city and county jail sample* (Doctoral dissertation). Retrieved from ProQuest Dissertation & Thesis Global (UMI No. 3220832).

Reason for rejection: Did not aggregate findings by gender

Lovell, D., Gagliardi, G. J., & Peterson, P. D. (2002). Recidivism and use of services among

persons with mental illness after release from prison. *Psychiatric Services*, 53(10), 1290-1297. doi: 10.1176/appi.ps.53.10.1290

Reason for rejection: did not aggregated findings by gender

Lovell, D., Gagliardi, G. J., & Phipps, P. (2005). *Washington's dangerous mentally ill offender law: Was community safety increased?* (Document No. 05-03-1901). Olympia, Washington: Washington State Institute for Public Policy.

Reason for rejection: Did not aggregate findings by gender

Loucks, A., & Zamble, E. (1999). Predictors of recidivism in serious female offenders: Canada searches for predictors common to both men and women. *Corrections Today*, 61(1), 26.

Reason for rejection: This article contained no data from which to calculate an effect size.

Maden, A. (1993). *Psychiatric disorder in women serving a prison sentence* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (ProQuest No. 10045769)

Reason for rejection: Did not conduct recidivism analyses

Maden, A., Friendship, C., McClintock, T., & Rutter, S. (1999). Outcome of admission to a medium secure psychiatric unit: I. Short- and long-term outcome. *British Journal of Psychiatry*, 175, 313-316.

Reason for rejection: Did not aggregate findings by gender; sample of women < 10.

Maden, A., Scott, F., Burnett, R., Lewis, G. H., Skapinaksis, P. (2004). Offending in psychiatric patients after discharge from medium secure units: prospective national cohort study. *British Medical Journal*, 328, 1534. doi: 10.1136/bmj.38075.467569.EE

Reason for rejection: Did not aggregate findings by gender; did not contain any codeable mental health variables

Maden, A., Skapinakis, P., Lewis, G., Scott, F., & Jamieson, E. (2006). Gender differences in reoffending after discharge from medium-secure units. National cohort study in England

and Wales. *The British Journal of Psychiatry: The Journal of Mental Science*, 189, 168-72. doi: 10.1192/bjp.bp.105.014613

Reason for rejection: This study had no non-mentally ill control group. Did not compare rates of recidivism between group of mentally ill offenders/group of offenders with a particular mental disorder/indicator and a group of offenders without the mental illness/indicator.

Mahoney, A. L., & Karatzias, T. (2012). Violent female offending: An exploration of repeat and one-time offending. *International Journal of Forensic Mental Health*, 11(3), 191-202. doi:10.1080/14999013.2012.723664

Reason for rejection: This was not a recidivism study. The authors investigated differences between women offenders who had committed one violent offense and those who had committed more than one violent offense (retrospectively). The temporal order of mental health problem/indicator/disorder and recidivism event could not be ascertained.

Marcoux, K. (2010). *Recidivism among mentally disordered offenders: An Ontario study* (Unpublished Master's thesis). University of Ottawa, Ottawa, Ontario, Canada.

Reason for rejection: There was no recidivism data by gender and mental disorder

Marshall, E. A., & Miller, H. A. (2019). Arbitrary decision making in the absence of evidence: An examination of factors related to treatment selection and recidivism for female sexual offenders. *Journal of Sexual Aggression*, doi: 10.1080/13552600.2019.1611961

Reason for rejection: Did not contain any mental health data.

McCarthy, L., Page, K., Baxter, H., Larkin, E., Cordess, C., & Duggan, C. (2001). Mentally disordered parricide and stranger killers admitted to high-security care 2: Course after release. *Journal of Forensic Psychiatry and Psychology*, 12(3), 501-514.

Reason for rejection: Did not aggregate findings by gender; samples of women too small

Mccooy, L.A., & Miller, H. A. (2013). Comparing gender across risk and recidivism in nonviolent offenders. *Women and Criminal Justice*, 23(2), 143-162. doi: 10.1080/08974454.2012.759054

Reason for rejection: No adequate mental health data.

Mcmurran, M., Egan, V., & Ahmadi, S. (1998). A retrospective evaluation of a therapeutic community for mentally disordered offenders. *The Journal of Forensic Psychiatry*, 9(1), 103-113. doi: 10.1080/09585189808402181

Reason for rejection: Did not aggregate findings by gender

Messina, N., Burdon, W., Hagopian, G., & Prendergast, M. (2004). One year return to custody rates among co-disordered offenders. *Behavioral Sciences and the Law*, 22, 503-518. doi: 10.1002/bsl.600

Reason for rejection: there were no mental health variables in recidivism analyses.

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., Roth, L. H., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York: Oxford University Press, Inc.

Reason for rejection: Did not aggregate by gender; author not contacted as additional data unlikely to be provided for this publication

Monson, C. M., Gunnin, D. D., Fogel, M. H., & Kyle, L. L. (2001). Stopping (or slowing) the revolving door: Factors related to NGRI acquittees' maintenance of a conditional release. *Law and Human Behavior*, 25(3), 257-267. doi: 10.1023/A:1010745927735

Reason for rejection: Did not aggregate findings by gender

Mundia, L., Matzin, R., Mahalle, S., Hamid, M. H.S., & Osman, R. S. (2017). Roles of psychopathic personality, mental health, and recidivism in criminal behavior: Survey of

Brunei inmates. *Asian Journal of Criminology*, 12(4), 255-280. doi: 10.1007/s11417-017-9249-6

Reason for rejection: Did not aggregate findings by gender; subsample of women <10

Muskens, M., Bogaerts, S., van Casteren, M., & Labrijn, S. (2011). Adult female sexual offending: A comparison between co-offenders and solo offenders in a Dutch sample. *Journal of Sexual Aggression*, 17(1), 46-60. doi: 10.1080/13552600.2010.544414

Reason for rejection: Did not examine impact of mental diagnoses on recidivism

Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing risk for violence among male and female civil psychiatric patients: The HCR-20, PCL: SV, and VSC. *Behavioral Sciences and the Law*, 22, 127-158. doi: 10.1002/bsl.579

Reason for rejection: Did not assess any mental disorders other than psychopathy

Odio, C.D., Carroll, M., Glass, S., Bauman, A., Taxman, F. S., & Meyer, J. P. (2018). Evaluating concurrent validity of criminal justice and clinical assessments among women on probation. *Health and Justice*, 6(7), 1-10. doi: 10.1186/s40352-018-0065-6

Reason for rejection: Only multivariate effects

O'Driscoll, C., Larney, S., Indig, D., & Basson, J. (2012). The impact of personality disorders, substance use and other mental illness on reoffending. *Forensic Psychiatry & Psychology*, 23(3), 382-391. doi: 10.1080/14789949.2012.686623

Reason for rejection: Did not aggregate findings by gender

Olson, D.E., Stalans, L. J., & Escobar, G. (2016). Comparing male and female prison releaseses across risk factors and postprison recidivism. *Women and Criminal Justice*, 26, 122-144. doi:10.1080/08974454.2015.1083930

Reason for rejection: Only multivariate effects reported

Ostermann, M., & Matejkowski, J. (2014a). Exploring the intersection of mental health and release status with recidivism. *Justice Quarterly*, *31*(4), 746-766, doi:

10.1080/07418825.2012.677465

Reason for rejection: Did not aggregate findings by gender

Ostermann, M., & Matejkowski, J. (2014b). Serious mental illness, criminal risk, parole supervision, and recidivism: Testing of conditional effects. *Law and Human Behavior*,

*39*(1), 75-56. doi: 10.1037/lhb0000094

Reason for rejection: Did not aggregate findings by gender.

Pantle, M. L., Pasewark, R. A., & Steadman, H. J. (1980). Comparing institutionalization periods and subsequent arrests of insanity acquittees and convicted felons. *The Journal of*

*Psychiatry and Law*, *8*, 305-316.

Reason for rejection: Sample of women at follow-up was too small ( $n < 10$ )

Pasewark, R. A., Pantle, M. L., & Steadman, H. J. (1982). Detention and rearrest rates of persons found not guilty by reason of insanity and convicted felons. *American Journal of*

*Psychiatry*, *139*(7), 892-897.

Reason for rejection: Sample of women was too small ( $n < 10$ )

Payne, C., McCabe, S., & Walker, N. (1974). Predicting offender-patients' reconvictions. *British Journal of Psychiatry*, *125*, 60-64.

Reason for rejection: Did not aggregate findings by gender; author contact info not found

Phillips, H. K., Gray, N. S., MacCulloch, S. I., Taylor, J., Moore, S. C., Huckle, P., & MacCulloch, M. J. (2005). Risk assessment in offenders with mental disorders - Relative

efficacy of personal demographic, criminal history, and clinical variables. *Journal of*

*Interpersonal Violence*, *20*(7), 833-847. doi: 10.1177/0886260504272898

Reason for rejection: Did not aggregate findings by gender.

Phillips, H. K., & Pasewark, R. A. (1980). Insanity plea in Connecticut. *Bulletin of the American Academy of Psychiatry and the Law*, 8(3), 335-344.

Reason for rejection: Contained no codeable data

Piper, A.J., & Nagy, V. M. (2018). Risk factors and pathways to imprisonment among incarcerated women in Victoria, 1860-1920. *Journal of Australian Studies*, 42(3), 268-284. doi:10.1080/14443058.2018.1489300

Reason for rejection: This study was excluded because the temporal order of “transfer to lunatic asylum” (which was the only mental health measure in that study) and recidivism could not be established.

Polvi, N. H. (1999). *The prediction of violence in pre-trial forensic patients: The relative efficacy of statistical versus clinical predictions of dangerousness* (Unpublished doctoral dissertation). Simon Fraser University, British Columbia, Canada.

Reason for rejection: Did not aggregate findings by gender

Putkonen, H., Komulainen, E. J., Virkkunen, M., Eronen, M., & Lönnqvist, J. (2003). Risk of repeat offending among violent female offenders with psychotic and personality disorders. *American Journal of Psychiatry*, 160(5), 947-951. doi: 10.1176/appi.ajp.160.5.947

Reason for rejection: Although the main sample consisted of mainly mentally ill women offenders, there was a non-negligible proportion that did not have any mental health diagnosis or had substance/alcohol abuse disorders. There was therefore no way to compare the mentally ill vs. non-mentally ill offenders in this study.

Quinsey, V. L., Jones, G. B., Book, A. S., & Barr, K. N. (2006). The dynamic prediction of antisocial behavior among forensic psychiatric patients A prospective field study. *Journal of Interpersonal Violence, 21*(12), 1539-1565. doi: 10.1177/0886260506294238

Reason for rejection: Did not aggregate findings by gender.

Regan, J. J., Alderson, A., & Regan, W. M. (2003). Psychiatric disorders in aging prisoners. *Clinical Gerontologist, 26*(1-2), 117-124. doi: 10.1300/J018v26n01\_10

Reason for rejection: Not a recidivism study

Rettinger, J. L., & Andrews, D.A. (2010). General risk and need, gender specificity, and recidivism of female offenders. *Criminal Justice and Behavior, 37*(1), 29-46. doi: 10.1177/0093854809349438

Reason for rejection: Did not measure mental health or relate it to recidivism.

Rezansoff, S. N., Moniruzzaman, Akm., Gress, C., & Somers, J. M. (2013). Psychiatric diagnoses and multiyear criminal recidivism in a Canadian provincial offender population. *Psychology, Public Policy, and Law, 19*(4), 443-453. doi: 10.1037/a0033907

Reason for rejection: This study did not aggregate by gender.

Rice, M. E., & Harris, G. T. (1992). A comparison of criminal recidivism among schizophrenic and nonschizophrenic offenders. *International Journal of Law and Psychiatry, 15*, 397-408.

Reason for rejection: All male sample

Rosemary, H., & Trotter, C. (2017). *Women's transitions from prison. The Post-release experience*. London: Routledge. doi: 10.4324/9781315455693

Reason for rejection: This book chapter contained no useable mental health data

Rowlands, R., Inch, H., Rodger, W., & Soliman, A. (1996). Diverted to where? What happens to the diverted mentally disordered offender. *The Journal of Forensic Psychiatry*, 7(2), 284-296.

Reason for rejection: Did not aggregate findings by gender

Scott, C. K., & Dennis, M. K. (2012). The first 90 days following release from jail: Findings from the Recovery Management Checkups for Women Offenders (RMCWO) experiment. *Drug and Alcohol Dependence*, 125(1-2)110-118. doi: 10.1016/j.drugalcdep.2012.03.025

Reason for rejection: No coefficients for the relationship between mental health variables and recidivism.

Serowik, K. L., & Yanos, P. (2013). The relationship between services and outcomes for a prison reentry population of those with severe mental illness. *Mental Health and Substance Use*, 6(1), 4-14. doi: 10.1080/17523281.2012.660979

Reason for rejection: No data for women offenders

Skeem, J. L., Eliza, W., Kennealy, P. J., Loudon, J. E., Tatar, J. R. (2014). Offenders with mental illness have criminogenic needs too: Toward recidivism reduction. *Law and Human Behavior*, 38(3), 212-224. doi: 10.1037/lhb0000054

Reason for rejection: Did not aggregate findings by gender.

Snowden, R. J., Gray, N., Taylor, J., & MacCulluch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine*, 37, 1539-1549. doi: 10.1017/S0033291707000876

Reason for rejection: All male sample

Solomon, P., & Draine, J. (1995). One-year outcomes of a randomized trial of case management with seriously mentally ill clients leaving jail. *Evaluation Review, 19*(3), 256-273.

Reason for rejection: All participants had mental health problems (no healthy controls)

Solomon, P., Draine, J., & Marcus, S. C. (2002). Predicting incarceration of clients of a psychiatric probation and parole service. *Psychiatric Services, 53*(1), 50-56. doi: 10.1176/appi.ps.53.1.50

Reason for rejection: Did not aggregate findings by gender.

Soyka, M., Graz, C., Bottlender, R., Dirschedl, P., & Schoech, H. (2007). Clinical correlates of later violence and criminal offences in schizophrenia. *Schizophrenia Research, 94*, 89-98. doi: 10.1016/j.schres.2007.03.027

Reason for rejection: Participants were not offenders. Only a small portion of the participants had committed a crime prior to receiving treatment

Soyka, M., Morhart-Klute, V., & Schoech, H. (2004). Delinquency and criminal offenses in former schizophrenic inpatients 7-12 years following discharge. *European Archives of Psychiatry and Clinical Neuroscience, 254*, 289-294. doi: 10.1007/s00406-004-0495-0

Reason for rejection: Participants were not offenders. Only a small portion of the participants had committed a crime prior to receiving treatment

Soyka, M., & Zingg, C. (2010). Association for Methodology and Documentation in Psychiatry Profiles predict later risk for criminal behavior and violent crimes in former inpatients with affective disorder. *Journal of Forensic Science, 55*(3).doi: 10.1111/j.1556-4029.2010.01354.x

Reason for rejection: Did not aggregate findings by gender

Stadtland, C., Kleindienst, N., Kröner, C., Eidt, M., & Nedopil, N. (2005) Psychopathic traits and risk of criminal recidivism in offenders with and without mental disorders. *International Journal of Forensic Mental Health*, 4(1), 89-97. doi:10.1080/14999013.2005.10471215

Reason for rejection: Did not aggregate findings by gender.

Stewart, L., Wilton, G., & Cousineau, C. (2012) *Federally Sentenced Offenders with Mental Disorders: Correctional Outcomes and Correctional Response* (Research Report R-268). Ottawa, ON: Correctional Service of Canada.

Reason for rejection: Did not aggregate by gender; authors not contacted because this was a CSC report

Swaminath, R. S., Mendonca, J. D., Vidal, C., & Chapman, P. (2002). Experiments in change: Pretrial diversion of offenders with mental illness. *Canadian Journal of Psychiatry*, 47(5), 450-458. doi: 10.1177/070674370204700506

Reason for rejection: Did not aggregate by gender; MH variables were not used to predict recidivism

Taylor, N. (2005). *An analysis of the effectiveness of Santa Clara County's Mentally Ill Offender Crime Reduction Program*. Retrieved from ProQuest Dissertation & Theses Global (UMI No. 3181686).

Reason for rejection: Did not aggregate findings by gender

Tengstrom, A. (2001). Long-term predictive validity of historical factors in two risk assessment instruments in a group of violent offenders with schizophrenia. *Nordic Journal of Psychiatry*, 55(4), 243-249. doi: 10.1080/080394801681019093

Reason for rejection: All male sample

Theurer, G., & Lovell, D. (2008). Recidivism of offenders with mental illness released from prison to an intensive community treatment program. *Journal of Offender Rehabilitation*, 47(4), 385-406. doi: 10.1080/10509670801995023

Reason for rejection: This study did not contain the data necessary to calculate an effect size (i.e., there were no appropriate group comparisons [e.g., mentally disordered vs. non-disordered] and no appropriate correlational data).

Thomson, L., Davidson, M., Brett, C., Steele, J., & Darjee, R. (2008). Risk assessment in forensic patients with schizophrenia: The predictive validity of actuarial scales and symptom severity for offending and violence over 8 – 10 Years. *International Journal of Forensic Mental Health*, 7(2), 173-189. doi:10.1080/14999013.2008.9914413

Reason for rejection: Did not aggregate findings by gender; subsamples of women would be too small for analyses by gender

Torgersen, L. J. (2013). *Factors related to recidivism among mentally disordered offenders: Differential impact of historical and dynamic clinical indicators across risk level* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Accession No.3602238)

Reason for rejection: Did not aggregate findings by gender

Trotter, C. (2017). 7 Return to reoffending\*. In R. Sheehan, & C. Trotter (Eds.), *Women's transitions from prison: The post-release experience* (pp. 103-112). London: Routledge.

Reason for rejection: The only mental health predictor was a mental health variable that combined several different diagnoses as well as various mental health predictors

Van Voorhis, P., Bauman, A., & Brushett, R. (2013). *Revalidation of the Women's Risk Needs Assessment: Institutional results-Final report*. Cincinnati, OH: Center for Criminal Justice Research.

Reason for rejection: Did not have a recidivism outcome variable

Wan, J. J., Morabito, D. J., Khaw, L., Knudson, M., & Dicker, R. A. (2006). Mental illness as an independent risk factor for unintentional injury and injury recidivism. *The Journal of Trauma, Injury, Infection, and Critical Care*, *61*(6), 1299-1304. doi: 10.1097/01.ta.0000240460.35245.la

Reason for rejection: Contained absolutely no relevant data. The term injury recidivism refers to psychiatric patients returning to hospital with a personal injury, and not to criminal recidivism

Warren, L. J., Mullen, P. E., Thomas, S. D., Ogloff, J. R., & Burgess, P. M. (2008). Threats to kill: a follow-up study. *Psychological Medicine*, *38*, 599-605. doi: 10.1017/S003329170700181X

Reason for rejection: Did not aggregate findings by gender

Wiederanders, M. R. (1992). Recidivism of disordered offenders who were conditionally vs. unconditionally released. *Behavioral Sciences and the Law*, *10*, 141-148.

Reason for rejection: No adequate non-mentally ill/different mental illness comparison group

Willging, C. E., Malcoe, L. H., Cyr, St. S., Zywiak, W. H., & Lapham, S. C. (2013). Behavioral health and social correlates of reincarceration among Hispanic, Native American, and White rural women. *Psychiatric Services*, *64*(6), 590-593. doi:10.1176/appi.ps.201200120

Reason for rejection: Cross-sectional design. Investigating correlates of previous incarcerations.

Wilson, A. B., Draine, J., Hadley, T., Metraux, S., & Evans, A. (2011). Examining the impact of mental illness and substance use on recidivism in a county jail. *International Journal of Law and Psychiatry*, 34, 264-268. doi: 10.1016/j.ijlp.2011.07.004

Reason for rejection: Did not aggregate findings by gender

Wilson, A. B., Draine, J., Barrenger, S., Hadley, T., & Evans Jr, A. (2014). Examining the impact of mental illness and substance use on time till re-incarceration in a county jail. *Administration and Policy of Mental Health and Mental Health Services*, 41, 293-301. doi: 10.1007/s10488-013-0467-7

Reason for rejection: Did not aggregate findings by gender

Wilson, J. A., & Wood, P. B. (2014). Dissecting the relationship between mental illness and return to reincarceration. *Journal of Criminal Justice*, 42, 527-537. doi: 10.1016/j.jcrimjus.2014.09.005

Reason for rejection: Did not aggregate findings by gender

Wilton, G., & Stewart, L. A. (2012). *Outcomes for offenders with concurrent substance abuse and mental health disorders* (Report No. R277). Correctional Service of Canada, Ottawa, Ontario, Canada.

Reason for rejection: Did not aggregate by gender; authors were not contacted because this was a government study

Wright, E.M., Van Voorhis, P., Bauman, A., & Salisbury, E. (2007). *Gender-responsive risk/needs assessment. Final Report*. [Prepared for the Minnesota Department of Corrections and the Advisory Task Force on the Woman and Juvenile Female Offender in Corrections]. Unpublished manuscript. Retrieved from University of Nevada, Las

Vegas website: [https://www.unlv.edu/sites/default/files/page\\_files/27/Minnesota-WRNA-Final-Report.pdf](https://www.unlv.edu/sites/default/files/page_files/27/Minnesota-WRNA-Final-Report.pdf)

Reason for rejection: No mental health variables included in this study

Yoskikawa, K., Taylor, P. J., Yamagami, A., Okada, T., Ando, K., Taruya, T., Matsumoto, T.,  
..Kikuchi, A. (2007). Violent recidivism among mentally disordered offenders in Japan.  
*Criminal Behaviour and Mental Health*, 17, 137-151. doi: 10.1002/cbm

Reason for rejection: Did not aggregate findings by gender

Zakaria, D., Allenby, K., Derkzen, D., & Jones, N. (2013). *Preliminary development of a dynamic risk assessment tool for women: An examination of gender neutral and gender specific variables* (Research Report R-280). Ottawa, ON: Correctional Service of Canada.

Reason for rejection: No useable mental health data.

\* Indicates studies which samples overlapped completely with an included study but from which only demographic/sample characteristics were gleaned to the exclusion of effect size data

## Appendix I

### Inter-rated Reliability Analyses: Exclusions and Strategic Approaches

A total of six variables initially intended for analyses were excluded post-coding. Reasons for exclusion were that a) none of the relevant studies reported the necessary data to allow for coding or b) the variables were wholly irrelevant to the studies themselves (e.g., (sub)sample data were irrelevant because no subsample was reported in the study).

#### Excluded variables

- Sample size after attrition
- Subsample sizes before and after attrition
- Mean age after attrition
- Attrition at follow-up
- Percentage of sample(s) that were treated for their mental health problem
- Percentage of sample(s) that were medication compliant
- Means and standard deviations (*SD*) of continuous scores on mental health predictors

#### Strategies for the Management of Artificial Inflation of Reliabilities

- 1) Variables that could not be coded on account of insufficient data were excluded from analyses; if “N/A” or “unknown” responses were included in analyses coders were likely to identify them with 100% accuracy, resulting in inflated ICC and Kappa values. Including this in estimates of inter-rater agreement would not reflect true comprehension of the variables nor to coders’ ability to adhere consistently to coding guidelines.
- 2) Some original variables used in the original dataset contained very large numbers of response options (e.g., > 40). In these cases, only the response options relevant to the four coded studies were entered in the datasets built for reliability analyses. Including all possible response options, when the overwhelming majority would be found by both coders to be obviously irrelevant, would again increase the risk of artificially inflating the values representing level of coder agreement.
- 3) ICC analyses in SPSS involves important data entry decisions. To avoid undue inflation in these analyses specifically, all relevant effect sizes reported in the original study were entered in separate rows, while study and sample characteristics included in the double-coding procedure were entered only alongside the first effect size listed (i.e., row 1). This strategy allowed for inter-rater agreement on study characteristics that were shared by, and identical across, multiple effect sizes, to be calculated only once. Naturally, for effect size-related variables each variable had to be coded separately for each effect size and all were included in analyses.

**Appendix J**

**SPIIn-W Pre-Screen Assessment (Orbis Partners, 2006)**

<b>A</b>	<b><i>Criminal History</i></b>			
1.	Age at first arrest	<i>Age at 1<sup>st</sup> Offenses</i>		
2.	Previous adult convictions	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2
3.	Incarcerations as an adult	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2
4.	Delinquency adjudications	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2
5.	Incarcerations as a delinquent	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2
6.	Variety of offenses	<input type="checkbox"/> Assault/Violence <input type="checkbox"/> Robbery <input type="checkbox"/> Burglary	<input type="checkbox"/> Fraud <input type="checkbox"/> Other Property <input type="checkbox"/> Drug	<input type="checkbox"/> DWI Offenses <input type="checkbox"/> Sex Offenses <input type="checkbox"/> Prostitution <input type="checkbox"/> Other

<b>B</b>	<b><i>Response to Supervision</i></b>		
1.	Failures to appear in court	<input type="checkbox"/> No	<input type="checkbox"/> Yes
2.	Technical Violations on Supervision	On Probation	<input type="checkbox"/> No
		On parole	<input type="checkbox"/> No
3.	New Offenses on Supervision	On Probation	<input type="checkbox"/> No
		On parole	<input type="checkbox"/> No
4.	Transfer to Custody while on Supervision	On Probation	<input type="checkbox"/> No
		On parole	<input type="checkbox"/> No
5.	Ever Absconded on Probation or Parole	<input type="checkbox"/> No	<input type="checkbox"/> Yes

<b>6.</b>	<b>Ever Escaped or Attempted Escape</b>			<input type="checkbox"/> <b>No</b>	<input type="checkbox"/> <b>Yes</b>	
<b>C</b> <i>Family and Children</i>						
<b>1.</b>	<b>Marital/Intimate Relationships</b>	<input type="checkbox"/> -- --	<input type="checkbox"/> --	<input type="checkbox"/> 0	<input type="checkbox"/> +	<input type="checkbox"/> ++
		High instability conflict, & dissatisfaction	Some conflict & dissatisfaction	Minimal satisfaction/no relationship	Stability & satisfaction	High stability, satisfaction & commitment
<b>2.</b>	<b>Relationship Risk Factors</b>	<input type="checkbox"/> Victim of domestic violence <input type="checkbox"/> Victimized: current/recent partner <input type="checkbox"/> Ongoing conflict: ex-partner <input type="checkbox"/> Safety/protection issues: spouse			<input type="checkbox"/> Perpetrated domestic violence <input type="checkbox"/> Partner with anti-social history <input type="checkbox"/> + Partner has pro-social influence <input type="checkbox"/> 0 N/A or none of the above	
<b>3.</b>	<b>Custody Arrangements</b>	<input type="checkbox"/> -- --	<input type="checkbox"/> --	<input type="checkbox"/> 0	<input type="checkbox"/> +	<input type="checkbox"/> ++
		Major dissatisfaction	Source of stress	n/a/ no issue	General satisfaction	High satisfaction

<b>D</b> <i>Social Network</i>			
<b>1.</b>	<b>Peer relationships</b>	<input type="checkbox"/> No (consistent) friends <input type="checkbox"/> 1+ friends antisocial influence <input type="checkbox"/> 1+ friends antisocial history <input type="checkbox"/> Only antisocial friends	<input type="checkbox"/> 1+ Friends prosocial influence <input type="checkbox"/> 1+ Close friends prosocial influence <input type="checkbox"/> 0 None of the above
<b>2.</b>	<b>Gang association</b>	<input type="checkbox"/> Belongs to gang <input type="checkbox"/> Gang member associates	<input type="checkbox"/> Family gang member <input type="checkbox"/> No gang associates

<b>E Substance Use</b>											
<b>Alcohol &amp; Drug Use</b>	<b>Ever Used</b>	<b>Times used before current conviction/last 3 months:</b>					<b>Age 1<sup>st</sup> Use (-)</b>	<b>Disrupts function (- -)</b>	<b>Contributes to criminal behaviour (- -)</b>	<b>Use in Custody (- -)</b>	<b>Attempts to cut back (+)</b>
<input type="checkbox"/> Alcohol/Drug Use <input type="checkbox"/> No Alcohol/Drug Use		<i>Daily (- -)</i>	<i>3/6 Days Weekly(-)</i>	<i>1-2 Days Weekly (-)</i>	<i>Fewer (0)</i>	<i>None (0)</i>					
Alcohol	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Marijuana	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Cocaine/Crack	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ecstasy/other club drugs	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Heroin	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Hallucinogens (LSD, Acid)	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Inhalants/huffing	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Amphetamines (Speed)	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Methamphetamine	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Prescription drug misuse	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Other	<input type="checkbox"/>						<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

<b>F</b> <i>Employment</i>						
<b>1.</b>	<b>Education</b>	<input type="checkbox"/> Less than 9 <sup>th</sup> grade <input type="checkbox"/> Less than 12 <sup>th</sup> grade <input type="checkbox"/> Literacy difficulties	<input type="checkbox"/> + High school graduate/GED <input type="checkbox"/> + Some post-secondary training	<input type="checkbox"/> + Trades program <input type="checkbox"/> + College degree <input type="checkbox"/> + Advanced degree		
<b>2.</b>	<b>Employment history</b>	<input type="checkbox"/> Unemployment at the time of current offense <input type="checkbox"/> Never employed more than 1+ year at a time <input type="checkbox"/> Never employed more than +6 months at a time <input type="checkbox"/> Frequently quit jobs (3+ times)	<input type="checkbox"/> Fired <input type="checkbox"/> Interpersonal conflicts with staff/employers <input type="checkbox"/> Difficulty finding employment/earn a living <input type="checkbox"/> 0 None of the above			
<b>3.</b>	<b>Motivation to join labor force</b>	<input type="checkbox"/> -- --	<input type="checkbox"/> --	<input type="checkbox"/> 0	<input type="checkbox"/> +	<input type="checkbox"/> ++
		No interest	Low	No issue	Desires employment	Intrinsically motivated/enjoys

<b>G</b> <i>Attitudes</i>						
<b>1.</b>	<b>Law-abiding attitudes</b>	<input type="checkbox"/> -- --	<input type="checkbox"/> --	<input type="checkbox"/> 0	<input type="checkbox"/> +	<input type="checkbox"/> ++
		Unwilling to demonstrate law-abiding behavior	Feels law abiding behavior does not apply to her	Not an issue	Desires to live in law-abiding manner	Commitment to law-abiding behavior
<b>2.</b>	<b>Attitudes toward the criminal justice system</b>	<input type="checkbox"/> -- --	<input type="checkbox"/> --	<input type="checkbox"/> 0	<input type="checkbox"/> +	<input type="checkbox"/> ++
		Contemptuous	Resentful	Not an issue	Appreciative	Respecting of

<b>H Social/Cognitive Skills</b>						
		<input type="checkbox"/> -- --	<input type="checkbox"/> --	<input type="checkbox"/> 0	<input type="checkbox"/> +	<input type="checkbox"/> ++
<b>1. Relationships</b>		Unable to form mutually rewarding relationships	Difficulty forming mutually rewarding relationships	Not an issue	Has mutually rewarding relationships	Capacity to form healthy mutually rewarding relationships
<b>2. Expression of needs</b>		Conflict/lack of clarity always present	Difficulty expressing effectively	Not an issue	Can express in assertive non-confrontational way sometimes	Expresses in assertive non-confrontational way

<b>I Mental Health</b>							
<b>1. Mental Health Conditions</b>	<i>Age of Onset</i>	<i>Past treatment</i>	<i>Current Condition</i>	<i>Treatment Received</i>	<i>Medication Prescribe</i>	<i>Compliance with Medication</i>	<i>Condition now Stable</i>
<input type="checkbox"/> No Conditions							
<input type="checkbox"/> Conditions							
Depression/Affective disorder							
Anxiety disorder							
Bi-polar Disorder							
Borderline Personality disorder							
Thought/adjustment Disorder							
PTSD							
Schizophrenia							
Other Psychoses							
Other							

<b>2. Abuse</b>	<i>None</i>	<i>As a Child</i>	<i>As an Adult</i>	<i>Currently</i>
Physical Abuse				
Sexual Abuse				
Emotional Abuse				

<b>4. Homicidal Ideation</b> (Attempts or thoughts to seriously harm others):	<input type="checkbox"/> 0 No Indicators <input type="checkbox"/> Indicators <input type="checkbox"/> None
---	--

<b>5. Suicidal Ideation</b> (Attempts or thoughts to seriously harm self):	<input type="checkbox"/> 0 No Indicators <input type="checkbox"/> - Suicidal thoughts <input type="checkbox"/> -- Suicidal attempts
--	---

<b>J</b>	<i>Violence</i>				
<b>1.</b>	<b>Violent behaviour in the last 6 months</b>	<input type="checkbox"/> No		<input type="checkbox"/> Yes	
<b>2.</b>	<b>Previous violent behaviour or convictions</b>	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3+
<b>3.</b>	<b>Any violence toward unknown victims</b>	<input type="checkbox"/> No		<input type="checkbox"/> Yes	
<b>4.</b>	<b>Perpetrators of domestic violence</b>	<input type="checkbox"/> No		<input type="checkbox"/> Yes	
<b>5.</b>	<b>Violations of protection or non-contact orders</b>	<input type="checkbox"/> No		<input type="checkbox"/> Yes	

**Appendix K**

**Adverse Childhood Experiences Scale (ACES): 12-Item SPIn-W Proxy Scale**

**Table 1K.**

*Items of the Adverse Childhood Experiences Scale (ACES): 12-Item SPIn-W Proxy Scale*

ACES Items	Domain	Item #
<i>Childhood abuse</i>		
Childhood sexual abuse	Domain I - MH	2a
Childhood physical abuse	Domain I - MH	2b
Childhood emotional abuse	Domain I - MH	2c
<i>Childhood neglect</i>		
Childhood emotional neglect	Domain C – F&C	7d
Childhood physical neglect	Domain C – F&C	7e
<i>Household Dysfunction</i>		
Household substance abuse	Domain C – F&C	7k
Household mental illness	Domain C – F&C	7l
Household member incarcerated	Domain C – F&C	7m
Mother treated violently	Domain C – F&C	7c
<i>Unstable home</i>		
Parental divorce/separation	Domain C – F&C	7b
Foster placements	Domain C – F&C	7j
Frequent conflict with parents	Domain C – F&C	7i

*Note.* MH = Mental Health; F&C = Family and Children.

## **Appendix L**

### **Additional LCA Models Tested**

In the process of examining the originally proposed models- Models 1 through 4 – three questions calling for additional empirical investigation arose. First, the question arose of whether including adverse childhood experiences as an auxiliary variable in the LCA models was adequately theoretically and empirically supported: both theoretical models and empirical research findings suggest that ACEs precede mental health problems. Thus, ACEs could be expected to have a role to play in the formation of the latent mental health classes (i.e., to have the causal effect of a covariate), rather than functioning solely as an additional class descriptive. Second, the most advantageous combination of an adequate sample size and an adequate base rate is not easily derived. Thus, models with longer follow-up times (i.e., higher base rates) but reduced samples were also tested. Third, methodologists have recommended that LCA indicators with cell frequencies less than five (or less than 10 in some cases) be removed from analyses, such that overall model fit and classification accuracy can improve. Although initial analyses included all possible data, even with lower than recommended cell frequencies, the impact of removing the three most infrequently endorsed indicators on model fit was also tested.

#### **ACEs: Auxiliary or Covariate?**

Briefly, while ACEs was originally included as an auxiliary variable, theoretical work and empirical evidence has long suggested that these early traumatic experiences contribute causally to adult mental health problems. Given then, that ACEs could, at least at the conceptual level, potentially constitute a predictor of class membership; it could be more suitable for inclusion as a covariate. Thus, a model was tested wherein ACEs was removed as an auxiliary variable and included alongside the Pre-Screen risk score as a covariate. Not surprisingly given

the absence of large class differences on this measure, this alteration in model structure failed to result in any meaningful changes in outcome, either in terms of model fit or in terms of class structure. Given this lack of support for ACEs as a significant contributor to initial class formation, this alternative model was rejected and ACEs remained an auxiliary variable in the main models.

### **Recidivism Alternatives: Sample Size versus Base Rates**

Second, the current author had initial concerns regarding the unusually low base rate of recidivism for the sample that were due to the restrictive recidivism definition including only returns to a Maine state prison. At 24 months follow-up, only 677 out of the original 920 cases had follow-up data, and less than 10% recidivated (refer to Table 1 in Methods, Participant section above). At 36 months, the number of valid cases dropped to 551, with a modest 5% gain in recidivism base rate. Rules of thumb, if not firmly established guidelines, exist with regards to base rates, particularly base rates for samples of justice-involved women. Generally, a base rate below 10% would be considered inadequate for recidivism analyses (Tabachnick & Fidell, 2019). Thus, given the inadequate base rate at 24 months follow-up, and despite the associated reduction in sample size, the 36-month recidivism alternative was also examined in a separate model. Perhaps not surprisingly, no improvements in the solution resulted from reliance on the longer follow-up time and a smaller sample size. Given the very large reduction in number of valid cases associated with the 48-month and 60-month recidivism outcomes, models were not tested with these outcomes.

### **Three Weak Indicators**

A careful inspection of the probability scale for each indicator across models and class-solutions resulted in the discovery of extremely low probability rates associated with three

mental health indicators – “Thought and adjustment Disorders”, “Somatization”, and “Homicidal Ideation”. It seemed plausible that these three items could negatively affect class separation and classification accuracy. Thus, both Model 2 and Model 4 with the favored three-class solution were re-examined with these three items excluded. The results indicated that a slight improvement in fit could be achieved by excluding these items. However, due to the very minute improvement associated with their exclusion, and the relative importance put on retaining as many indicators of mental health as possible and thereby supporting the most complete understanding of the nature of the classes, the decision was made to keep all the original items in the main models, including the three low frequency items; the reduced 11-item models were rejected.

Appendix M

Memorandum of Understanding (MOU)

MAINE DEPARTMENT OF CORRECTIONS

I have read the Department of Corrections Policy 1.24, Research, Evaluation, and Performance Measurement, and agree to comply with it.

Signed: [Redacted] Date: 12.4.17

Title: Dr. David Robinson, President & CEO, Orbis Partners Inc.

Signed: [Redacted] Date: 12-04-17

Title: Dr. Marilyn Van Dieren, Senior Partner, Orbis Partners Inc.

Signed: [Redacted] Date: 12 Oct. 17

Title: Dan Abet, Analyst, Orbis Partners Inc.

Signed: [Redacted] Date: 12/4/17

Title: Erica King, Senior Associate, Orbis Partners, Inc.

Signed: [Redacted] Date: 02/05/19

Title: Dr. Shelley Brown, Associate Professor, Dept. Psychology, Carleton University

Signed: [Redacted] Date: 02/06/19

Title: Cathrine Pettersen, Doctoral Candidate, Dept. Psychology, Carleton University

The above noted research project has been reviewed and approved in accordance with Department of Corrections Policy 1.24 Research, Evaluation, and Performance Measurement

Director of Correctional Program Practices or designee Date

## Appendix N

## Certification of Institutional Ethics clearance



Office of Research Ethics  
4500 ARISE Building | 1125 Colonel By Drive  
Ottawa, Ontario K1S 5B6  
613-520-2600 Ext: 4085  
[ethics@carleton.ca](mailto:ethics@carleton.ca)

**CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE**

The Carleton University Research Ethics Board-B (CUREB-B) has granted ethics clearance for the research project described below and research may now proceed. CUREB-B is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).

**Ethics Protocol Clearance ID:** Project # 112238

**Research Team:** Ms. Cathrine Petterson (Primary Investigator)  
Dr. Shelley Brown (Research Supervisor)

**Project Title:** Mental Health Disorders and Recidivism Among Adult Women Offenders

**Funding Source** (if applicable):

**Effective:** January 31, 2020

**Expires:** January 31, 2021.

**Please ensure the study clearance number is prominently placed in all recruitment and consent materials: CUREB-B Clearance # 112238.**

**Restrictions:**

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.
2. Any modification to the approved research must be submitted to CUREB-B via a Change to Protocol Form. All changes must be cleared prior to the continuance of the research.
3. An Annual Status Report for the renewal of ethics clearance must be submitted and cleared by the renewal date listed above. Failure to submit the Annual Status Report will result in the closure of the file. If funding is associated, funds will be frozen.
4. A closure request must be sent to CUREB-B when the research is complete or terminated.
5. During the course of the study, if you encounter an adverse event, material incidental finding, protocol deviation or other unanticipated problem, you must complete and submit a Report of

**Appendix N****Certification of Institutional Ethics clearance**

Adverse Events and Unanticipated Problems Form, found here:  
<https://carleton.ca/researchethics/forms-and-templates/>

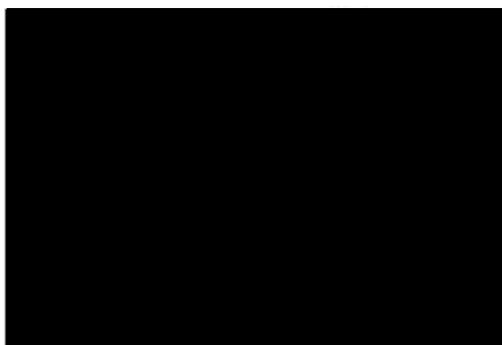
Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2nd edition* and the *Carleton University Policies and Procedures for the Ethical Conduct of Research* may result in the suspension or termination of the research project.

Upon reasonable request, it is the policy of CUREB, for cleared protocols, to release the name of the PI, the title of the project, and the date of clearance and any renewal(s).

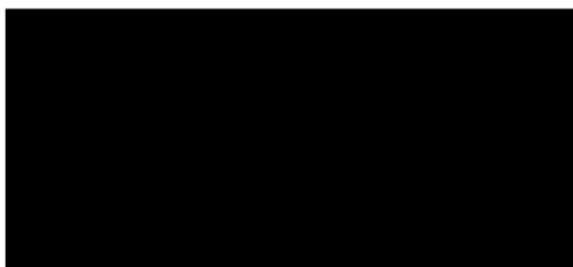
Please contact the Research Compliance Coordinators, at [ethics@carleton.ca](mailto:ethics@carleton.ca), if you have any questions.

**CLEARED BY:**

**Date: January 31, 2020**



Natasha Artemeva, PhD, Chair, CUREB-B



Janet Mantler, PhD, Vice-Chair, CUREB-B

Appendix O

Spin-W Mental Health Items: Indicator, Scaling, Risk Contributions, & LCA utility

Table 10.

SPIn-W Mental Health Items: Response Scale, Static Risk Score, LCA Use

Item #	Item Name	Response Scale	SR	LCA Use
I01A1	Mental health condition ( <i>none or conditions</i> )	0-1	5	<i>Excluded</i>
I01B1c	Depression/affective disorder - current	0-1	x	Indicator
I01B2c	Anxiety disorder - current	0-1	x	Indicator
I01B3c	Bipolar disorder - current	0-1	x	Indicator
I01B4c	BDP - current	0-1	x	Indicator
I01B5c	Thought/adjustment Disorders - current	0-1	x	Indicator
I01B6c	PTSD - current	0-1	x	Indicator
I01B7c&8c	Schizophrenia & other psychoses – current <sup>a</sup>	0-1	x	Indicator
I01B9c	Other: ADHD – current <sup>b</sup>	0-1	x	Indicator
I02A2	Physical abuse as a child	0-1	x	Auxiliary: ACEs
I02B2	Sexual abuse as a child	0-1	x	Auxiliary: ACEs
I02C2	Emotional abuse as a child	0-1	x	Auxiliary: ACEs
I03A	Self-injurious behaviour	0-1	x	Indicator
I03B	Eating disorders	0-1	x	Indicator
I03C	Somatization	0-1	x	Indicator
I04	Homicidal ideation	0-1	10	Indicator
I05	Suicidal ideation	0-2	10	Indicator
I06	Motivation to address mental health risk	0-4	x	<i>Excluded</i>
I07	Resources for mental health risk	0-4	x	<i>Excluded</i>

Note. Items referred to as Indicators were utilized in the LCA as were Auxiliary items. SR = Corresponding Static Risk Score. For all Items 1B the sub-item “current” was used to identify cases rather than the main Item 1Ba, as the latter encompassed both past and current disorders. Item 1B7c (*Schizophrenia*) and 1B8c (*Other Psychoses*) were combined into a single “Any Psychosis” indicator on account of both items reflecting psychoses and the very low prevalence rates expected for these disorders. <sup>b</sup> Item 1B9c (*Other disorder*) did not indicate the nature of the other unaccounted for disorder present, so the supporting text variable Item 1B9text that specifies the exact disorder was used to identify any additional disorders prevalent enough to constitute its own indicator. ADHD was the only disorder with adequate frequency rates. <sup>c</sup> As Item 3a alone was subject to occasional user errors wherein the assessor failed to identify this mental health problem at the time of the original assessment, Items 3d (*Other*) and 3dText (a text variable identifying the specific mental health marker in the “other” category) were examined to retrieve the cases wherein self-harm had only been identified within these variables.

## Appendix P

### Mplus Syntax: The Final Model 4: Three-Class Solution

! AUTHOR: C. PETTERSEN!  
 ! MPLUS SYNTAX – DISSERTATION STUDY 2: SPIN-W MENTAL HEALTH LCA!  
 ! DATE: 01-21-2021!  
 ! MODEL: 4 – 14 MENTAL HEALTH INDICATORS, PRE-SCREEN RISK SCORE COVARIATE!  
 ! THREE AUXILIARY VARIABLES: ACES, AGE, & RACE, ONE DISTAL OUTCOME: 24!  
 ! MONTHS RECIDIVISM!  
 ! CLASS SOLUTION: THREE CLASSES!

**TITLE:** M4\_C3\_R24;

**DATA:**

File is "C:\Users\cathrinepettersen.PCLAA405-03\Desktop\MH\_LCA\_M+\_FINAL.dat";

**FORMAT IS FREE;**

**VARIABLE:**

**NAMES ARE**

ID MOODCUR ANXCUR BIPCUR BPDCUR  
 THADCUR PTSDCUR PSYCUR ADHDCUR  
 DI3B DI3C DI3AALT DI4 ALTSU2 ACES  
 COMORBID PSCTOT AGE RACE REC24 REC36  
 REC48 REC60 ANYREC24 RETCUS;

**USEVAR ARE**

MOODCUR ANXCUR BIPCUR BPDCUR  
 THADCUR PTSDCUR PSYCUR ADHDCUR  
 DI3B DI3C DI3AALT DI4 ALTSU2  
 ACES COMORBID PSCTOT  
 AGE RACE REC24;

**CATEGORICAL ARE**

MOODCUR ANXCUR BIPCUR BPDCUR  
 THADCUR PTSDCUR PSYCUR ADHDCUR  
 DI3B DI3C DI3AALT DI4 ALTSU2  
 COMORBID;

**MISSING ARE ALL (999);**

**AUXILIARY = ACES AGE RACE (e);**

**CLASSES = c (3);**

**Appendix P (Cont.)****Mplus Syntax: The Final Model 4: Three-Class Solution (Cont.)****ANALYSIS:**

**ALGORITHM** = INTEGRATION;  
**TYPE** = Mixture;  
**STARTS** = 1000 250;  
**ESTIMATOR** = MLR;

**MODEL:** %Overall%

C on PSCTOT;

**OUTPUT:** Entropy Sampstat Tech10 Tech11 STDY;

**PLOT:**

**TYPE** = Plot3;  
**SERIES** = MOODCUR (1) ANXCUR (2) BIPCUR (3)  
BPDCUR (4) THADCUR (5) PTSDCUR (6)  
PSYCUR (7) ADHDCUR (8) DI3B (9) DI3C (10)  
DI3AALT (11) DI4 (12) ALTSU2 (13) COMORBID (14);

**SAVEDATA:** File is M4\_3C\_R24\_SPSS.dat;

**SAVE IS** CPROB;

**! SAVEDATA:** File is "C:\Users\cathrinepettersen.PCLAA405-03\Desktop\SPSS\M4\_3C\_R24\SPSS"!