EXAMINING THE EFFECTIVENESS OF INTERVENTIONS FOR CRIMINAL JUSTICE-INVOLVED WOMEN

A Meta-Analytic Review

RICHARD SUMMERS
SIMON PEMBERTON
JOANNA LONG
University of Birmingham

Academic understanding of women's gendered pathways into the criminal justice system has grown significantly over the last 20 years. Allied to this development has been an increasing number of gender-responsive practices and interventions designed to address the needs of criminal justice-involved women. This meta-analysis summarizes the trends in 71 interventions extracted from 64 papers involving justice-involved women. Subgroup analysis and meta-regression were used, which shows that gender-responsive interventions are up to 42% more effective compared with gender-neutral, even when controlling for several covariates. Other findings in our case demonstrate features of interventions, such as intervention, format, focus, and length, that appear to be effective in reducing recidivism for criminal justice-involved women. Our findings strengthen the case for investment in gender-responsive interventions and diversion programs.

Keywords: gender differences; meta-analysis; intervention; women offenders; gender

INTRODUCTION

The question of effective interventions for criminal justice-involved (CJI) women rose to prominence in both academic and policy circles from the end of the 1990s. This issue was prompted by rising female prison populations that outstripped the growth in the male prison estate in many advanced industrialized countries and coalesced with the emergence of the "what works" agenda in criminal justice policy. Although some writers attempted to adapt male-centric models to explain and respond to female offending others sought to develop gender-specific approaches. The work of Bloom et al. (2003) offered a framework of

AUTHORS' NOTE: *The authors thank the contribution of the anonymous reviewers and editors to this article. There are no conflicts of interest to disclose. The research on which this article was based was funded by The JABBS Foundation. Correspondence concerning this article should be addressed to Simon Pemberton, Department of Social Policy, Sociology and Criminology, University of Birmingham, Muirhead Tower, Edgbaston, Birmingham B15 2TT, UK; e-mail: s.pemberton.1@bham.ac.uk.*

CRIMINAL JUSTICE AND BEHAVIOR, 202X, Vol. XX, No. X, Month 2025, 1–26.
DOI: 10.1177/00938548241304753
Article reuse guidelines: sagepub.com/journals-permissions

principles that has since influenced the delivery of a growing number of gender-responsive services/interventions for CJI women. Running alongside these policy and practice developments is an evolving evidence base. As the work by Van Voorhis (2022) suggests, we now have a greater understanding of (a) the distinct-gendered pathways that women take into the criminal justice system (Brennan et al., 2012; Brennan & Jackson, 2022); (b) the specific gendered needs that, if unaddressed, are predictive of criminal justice involvement (Van Voorhis, 2012); (c) the interventions and case management practices that produce optimal results for CJI women (Gobeil et al., 2016). This article focuses on the later point and seeks to evaluate "what works" for CJI women.

The existing literature suggests that interventions which target either "criminogenic needs" or "gendered needs" of CJI women demonstrate a greater impact on recidivism when compared with standard treatment (Prison and Probationary Supervision). In short, doing something appears to be more effective than the status quo (Gobeil et al., 2016). Perhaps then, a more relevant question that this article seeks to address is "what might work best" for CJI women? To some extent, the meta-analysis undertaken in the work by Gobeil et al. (2016) provides some insights into this question—offering statistically tentative support for the contention that gender-responsive approaches are more effective than gender-neutral interventions. Beyond these headline findings, important gaps remain in our understanding. Specifically, a deeper understanding of why and for whom these interventions work is critical to developing our knowledge of what works best for women.

WHAT IS A GENDER-RESPONSIVE INTERVENTION?

Gender-responsive approaches recognize women's distinct needs, psychological development, and life experiences. Interventions should be designed specifically with these differences in mind. Consequently, Bloom et al. (2003) suggest three related theoretical frames that should inform gender-responsive practice. First, *Relational Theory* posits the importance of relationships specifically to women and the sense of self that develops through social connections made with others (Covington, 1998). Second, a *Theory of Addiction* which recognizes that women use substances as a means to make or sustain connections to others, or, to mask the pain of non-mutual, abusive relationships (Covington & Surrey, 1997). Third, a *Theory of Trauma* which views trauma as a disease of disconnection, and therefore, recovery necessarily involves creating safe spaces whereby women might heal through connection to others (Herman, 1992). In addition, we would add a fourth dimension, a *Theory of socioeconomic marginalization*, necessary to understand the material forms of disconnection CJI women experience that significantly frame and determine their life chances (Morash & Kashy, 2022). Specifically, this would acknowledge the gendered patterning of economic resources and opportunities, and the disproportionate economic/emotional burden that socially ascribed roles place onto women. Following the principles set in the work by Bloom et al. (2003) and others, a number of gender-responsive programs have developed in the intervening years. Drawing on these interventions, we identify four key components of established gender-responsive practice.

1. Gender-Responsive Interventions Must Primarily Address Women's Pathways into the Criminal Justice System

It is widely believed that women offend for different reasons to men. The literature suggests three gender-specific pathways which shape entry into the criminal justice system: the

Victimization pathway highlights the impact of childhood abuse on adult experience of poor mental health (post-traumatic stress disorder [PTSD], depression, or anxiety) and substance use issues that act as drivers of criminal justice involvement (Daly, 1992); the *Relational pathway* documents the interrelationship between dysfunctional and damaging intimate relationships which renders women vulnerable to patterns of co-offending with abusive partners (Covington, 1998); *the Deep marginalization pathway identifies* a pattern of poverty, homelessness, and social exclusion which means that some women offend due to a combination of economic necessity and the opportunities provided through their proximity to criminogenic networks (Richie, 2001, 2018). Gender-responsive interventions should seek to disrupt these pathways with programs that simultaneously address constellations of interrelated needs.

A Gendered Assessment of Needs and Strengths Should Inform Individualized Treatment/Service Provision

The personal histories of CJI women exhibit key differences to their male counterparts in terms of substance use, trauma, mental illness, parenting stress, employment histories, and housing situations. These histories and the needs that arise from them extend beyond the "Central 8" criminogenic needs (Andrews & Bonta, 2010) and consequently, traditional assessment tools fail to capture the gendered needs of women. Hence, gendered assessment tools, such as the Women's Risk Needs Assessment, are critical to the accurate and comprehensive identification of service users "needs," so that appropriate holistic service provision may be determined (Van Voorhis et al., 2010). Alongside needs, gendered assessments should identify the many "strengths" that CJI women exhibit and the positive relationships they draw on to navigate the challenges in their daily lives—so that these may be built on as known protective factors against the drivers of reoffending (Van Voorhis et al., 2010). Moreover, most CJI women have had exposure to controlling relationships and social marginalization, and consequently experience lower levels of self-worth; meaning that motivational interviewing is a critical feature of assessment/casework approaches to rebuild self-esteem (Miller & Rollnick, 2012).

Delivery Methods and Goals of Gender-Responsive Interventions

Given that many CJI women present with complex needs, gender-responsive interventions should be sensitive to dual or multiple treatment needs, addressing several related needs simultaneously. Gender-responsive interventions should be multimodal in nature using a range of methods and practices to holistically address multiple needs (Bloom et al., 2003). They should seek to address the "whole person," not just isolated cognitive processes, and specifically in the case of women, should be minded toward the need to be positively connected to others. Services should be both "designed by" and "designed for" the women who use them to ensure their engagement and to logically enhance the potential for their success (Bloom et al., 2003). Finally, gender-responsive interventions should promote women's journeys toward an ultimate goal of critical autonomy (Bloom et al., 2003). Thus, programs should aim to develop vocational and life skills, and to provide access to resources and opportunities, so that CJI women might eventually lead healthier and more fulfilling lives of their own choosing (Van Voorhis et al., 2010).

An Environment Designed for a Gender-Responsive Intervention

Gender-responsive interventions must take place within a physically and psychologically safe space, where service users are treated with respect and do not feel judged by those supporting them. Importantly, this space should not reflect the conditions or reproduce the emotions associated with control and abuse found within women's personal histories. Thus, these spaces should be women only and should have staff and mentors with whom the women can relate to in terms of lived experience and background. Critical to creating such an environment is the extent to which interventions are trauma-informed, in so far as staff can identify and understand the consequences of trauma and potential triggers that might re-traumatize. Treatment and healing are more likely to occur within safe, nurturing, and consistent environments.

"WHAT WORKS" FOR CJI WOMEN

It is well established in the extant literature that an intervention that is designed to address either identifiable "criminogenic needs" or broader "gendered needs" is preferable to treatment as usual (standard prison and probationary supervision) (Gobeil et al., 2016). That said, where this consensus unravels is the extent to which interventions that might be categorized as either "gender-neutral" or "gender-responsive" perform in relation to one another in reducing reoffending for CJI women. This speaks to a broader set of debates over the relevance of gendered pathways and distinct gendered needs/strengths to the design of interventions to address female offending. In particular, these debates have focused on the appropriateness of the Risk Needs Responsivity (gender-neutral) model to the experiences and life histories of CJI women.

GENDER-NEUTRAL INTERVENTIONS—RISK NEEDS RESPONSIVITY MODEL

In this section, we review the core principles of the Risk Needs Responsivity (RNR) Model as a key influence on the design of gender-neutral interventions. Taking elements each in turn, the Risk principle suggests that the form and intensity of interventions should be designed to address a sliding scale of risk. Thus, as the offending risk increases, the treatment dosage should be raised incrementally. There is an intuitive appeal to the risk principle; however, whether this is appropriate for CJI women requires further consideration. The fact that CJI women tend to present with multiple needs while demonstrating disproportionately lower levels of risk in terms of criminal harms, somewhat problematizes this logic. Moreover, as Messina and Esparza (2022) note that intervention dosage (length and intensity) may be less significant than is assumed by Risk Need Responsivity, rather more critical factors are the appropriateness of the content of the intervention; and the applicability to the needs of the target population.

This leads neatly onto the most debated aspect of RNR in relation to women, which is the Need principle. According to this principle, reductions in recidivism can only be achieved by addressing dynamic risk factors that are probabilistically linked to future offenses. According to the work by Andrews and Bonta (2010), eight central dynamic needs/risk factors drive criminal behavior, which comprised the "Moderate 4" indirect factors (poor/family relationships, lack of education/employment, lack of prosocial leisure time, and substance use) and the "Big 4" (antisocial personality, antisocial cognition, antisocial behavior, and antisocial associates) as direct predictors of recidivism. The "Central 8" needs/risks are

purported to be consistent across age, ethnicity, and gender; however, several critiques suggest that the "Central 8" are inappropriately generalized to women (Blanchette & Brown, 2006). Moreover, as Salisbury et al. (2016) suggest, even where the "Central 8" might be effective predictors for women of recidivism, validity is significantly improved through the inclusion of gender-responsive needs. Factors that do not feature in the "Central 8," particularly static factors, such as childhood and adulthood physical and sexual abuse, play a key contextual role in offending behaviors. Consequently, the RNR needs principle fails to account for the critical relationship between criminogenic needs and, as Messina and Esparza (2022) put it, "destabilizing factors" common to the lives of many CJI women. As Brennan et al.'s (2012) work suggests, for the few CJI women who exhibit characteristics aligned to the "Central 8," it is likely that gender-neutral interventions are effective; however, for the vast majority of CJI women who present with different needs/risk profiles to men, these interventions are unlikely to be optimal.

The Responsivity principle asserts that intervention formats should be matched to the learning styles, capacities, and characteristics of the individual who is subject to treatment (Andrews & Bonta, 2010). Yet, the contextual influence of gender is often lost here, and in particular, the role of trauma and the relational nature of substance use and violence in women's lives (Messina & Esparza, 2022). Ultimately, these factors for many CJI women will determine the effectiveness of intervention design for them. Given these issues, Belisle et al. (2022, p. 3) correctly argue that the "what works" agenda might have brought improvements to the policy in this area; however, there remains "a concern that when working with justice-involved women, the field frequently settles for 'what works' instead of 'what works best'." We begin to explore this claim in detail in the following section.

GENDER-RESPONSIVE INTERVENTIONS

As more interventions came to be informed by the gender-responsive principles set out in Bloom et al.'s work, an evidence base has developed through a growing number of evaluation studies, randomized control, and match group trials. Van Voorhis (2022, p. 139) has recently suggested that the evolution of the gender-responsive approach and practices has led to "a large body of research [which] is now offering an impressive empirical picture of 'What Works' [for CJI women]."

A number of multimodal, gender-responsive programs combine CBT with different therapies (mindfulness, physical exercise, art therapy) or skills training (budgeting, educational qualifications) to address specific gendered needs and to build related strengths. Positive results have been documented in three areas: (a) gender-responsive trauma programs (such as Helping Women Recover, Beyond Trauma, and Beyond Violence) have reported positive impacts on recidivism and secondary outcome measures (e.g., anxiety, depression, substance use, PTSD) (Messina et al., 2010); (b) evaluations of employment skills programs' Moving On' and "Girls Moving On," report favorable results and in particular decreases in recidivism (Gehring et al., 2010); and (c) parenting programs, such as "Parenting Inside Out" and "Emotions" demonstrate these programs successfully support women's relationships with their children, improve emotional regulation, and reduce recidivism (Eddy et al., 2013).

Gender-responsive approaches have also informed the development of case management practices, and specifically, probation models (Belisle et al., 2022). Advocates suggest that case management practices based on the relational principles of trust, caring, and fairness that run counter to the controlling and abusive relationships that have marked the women's

personal histories are more effective in reducing reoffending than traditional forms of supervision that prioritize toughness, threats, and surveillance (Skeem et al., 2009). Moreover, gender-responsive case plans that are coproduced with women and address multiple needs in a staged approach appear to be more effective in reducing recidivism than standard probation approaches (Millson & Van Dieten, 2010; Morash et al., 2016).

Finally, it is notable up to this point that most gender-responsive interventions are delivered in prison or at the point of release into the community. Whether custody or probation supervision is conducive to the delivery of trauma-informed services is an important question to pose. In some respects, the development of gender-responsive diversion programs, either at the point of arrest or the pre-sentence stage, provides a more appropriate environment for CJI women's recovery. Implicitly, these initiatives recognize that prolonged engagement with the system and specifically imprisonment can be counterproductive in terms of recidivism. In theory, diverting CJI women away from the system into gender-responsive services that address the contextual needs that drive female offending should be a more effective and less costly intervention than prison. The evidence base broadly supports this point (Myer & Buchholz, 2018), although it should be noted that a couple of these studies are not able to produce statistically significant reoffending outcomes due to low sample size (Brennan et al., 2012; Messina et al., 2012).

The growth in this evidence base led to the important meta-analysis undertaken by Gobeil et al. (2016). Drawing on 37 studies (38 effect sizes), Gobeil et al. (2016) investigated the effectiveness of interventions for CJI women and reached the following conclusions. First, any intervention (gender-neutral or gender-responsive) appears to work more effectively than no treatment at all. Second, when gender-responsive and gender-neutral interventions are compared, only high-quality studies detected a larger effect size for gender-responsive interventions. Gobeil et al. (2016) were unable to control for participant characteristics and therefore, cannot comment on for whom gender-responsive interventions might work. Third, in terms of the type of interventions most likely to reduce recidivism, substance abuse interventions ($k = 16$) produced the strongest effects. Interventions using therapeutic communities similarly produced strong effects. However, as Gobeil et al. (2016) note, 7 out of the 10 therapeutic community interventions also targeted substance use; therefore, it is difficult to disentangle the effects of these factors. Finally, the results of Gobeil et al. (2016) suggest that interventions conducted in either institutions or that operate between institutions and the community seem more effective than those conducted solely in the community.

This meta-analysis extends Gobeil et al.'s (2016) review and includes studies that have been published in the intervening years, to address the following aims: (a) to explore the extent to which gender-responsive interventions effectively reduce recidivism compared with gender-neutral interventions and (b) to investigate the specific features of interventions and contextual factors that influence their efficacy.

METHOD

DATABASE SEARCHES

Databases were chosen for their comprehensive coverage of disciplines and key journals most relevant to the research questions. Databases included PsycINFO, covering journals in the fields of sociology and law; Social Sciences Full Text encompassing areas, such as addiction studies and social work; Sociological Abstracts, focusing on international social and

behavioral sciences literature; Web of Science, which catalogs a wide range of scientific and social sciences subjects; and Criminal Justice Abstracts, specializing in criminal justice research. In addition, gray literature was sourced from OATD.org and the Bielefeld Academic Search Engine (BASE), which includes theses and other unpublished materials. The search terms were aligned with the work by Gobeil et al. (2016), which defined gender ("women," "woman," "female"), population ("offen*," "crim*," "priso*," "incarcerat*," "inmate," "detain*"), and intervention ("program*," "interven*," "treat*," "therap*," "rehab*"). The search syntax was adjusted according to the requirements of each database. Searches were limited to English and dates ranging from 1990 to 2024. This date range was selected to cover the emergence of the "what works" agenda in the early 1990s and to encompass the development of the gender-responsive approach/practice from the late 1990s.

STUDY SCREENING AND CODING

Following the process outlined in the PRISMA flowchart (Figure 1): In Step 1, the databases were searched in September 2023 by R.S., resulting in 61,173 records. These records were then exported, and subsequently, abstracts were shifted using Abstrackr (Wallace et al., 2012). After removing duplicated and completely irrelevant results, this left 14,806 abstracts to be screened (Step 2). In the eligibility-assessment stage (Step 3), clear inclusion/exclusion criteria were defined using the PICO guidelines (Mattos & Ruellas, 2015): Participants (CJI women); **I**ntervention/Exposure (Gender-responsive interventions); **C**ontrol (Treatment as Usual/Other Intervention); **O**utcome (Levels of criminogenic needs/ Recidivism/ Reconviction). Consequently, 14,333 studies were excluded because they did not match the PICO criteria (e.g., not a test of an intervention, no recidivism outcome etc.). This process left $k = 473$ studies for further assessment. R.S. screened all abstracts, and a 20% sample was also reviewed by S.P. and J.L. Details of classifications were discussed and agreed in the protocol, which was then used as the criteria by which to classify the papers (DOI 10.17605/OSF.IO/JQP52). There was a high agreement between reviewers (Kappa = .98) regarding the inclusion/exclusion of studies. Following this process, studies identified by Gobeil et al. (2016) were added, leaving a pool of 163 studies. As a final stage (Step 4), additional data were requested from 94 study authors, as study results could not be disaggregated for female participants in many instances. As a result of this process, 64 studies were included in the meta-analysis.

DATA EXTRACTION AND STUDY CLASSIFICATION

The data were extracted from each paper into a dataset, consisting of study details, such as the authors, sample size for experimental and control groups, recidivism events for experimental and control groups, length of follow-up (months), intervention length (months), and whether it was a female-only intervention.

Each author independently classified all eligible studies according to the study focus, location, delivery approach, gender-responsiveness, and SMS score (Maryland Scientific Methods Scale; Sherman et al., 1998). The agreement between authors (interrater reliability) was high; Kappa was calculated for each variable: inclusion = .89, focus = .86, location = .92, delivery approach = .92, Gender-responsive = .93, and SMS = .79.

The criteria by which these classifications were agreed upon were published in the protocol beforehand and described in detail below.

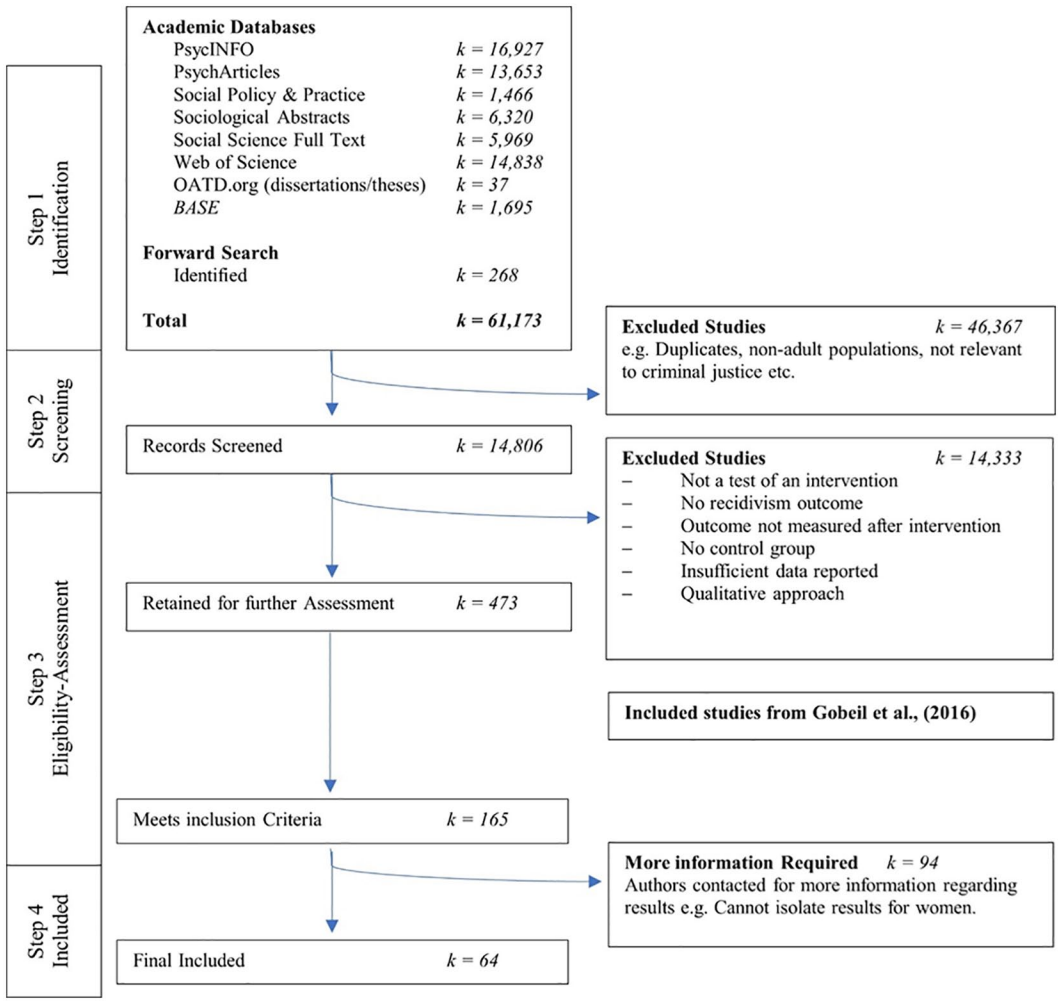

Figure 1 PRISMA flowchart, Adopted From Moher et al. (2009)
Note. PRISMA flowchart illustrating the systematic search and selection process for included studies in the meta-analysis.

Gender-Responsive Intervention (Gendered, Partially Gendered, and Gender-Neutral)

Studies were classed as gender-responsive, if they contained three or four of the criteria below (i.e., pathways, assessment, content, context). Studies were classed as partially gender-responsive, if they contained one or two of the gender-responsive criteria below. Studies that contained none of the gendered criteria were classified as gender-neutral:

1. **Gendered Pathways**—Gender-responsive interventions should address the aspects of any of the three pathways identified earlier within this article (and the clusters of needs that constitute these pathways), namely (a) the victimization pathway, (b) the relational pathway, and (c) the social marginalization pathway.
2. **Gendered Assessment**—A gender-responsive intervention should be based on an assessment that identifies an individual's needs and strengths, and uses a motivational interviewing

technique. The assessment must contain gender-specific scales, that is, abuse, trauma, parenting, and housing.

3. **Model and Aims of Delivery (content)**—A gender-responsive intervention should be delivered in the following form: (a) multimodal, (b) address dual or multiple treatment needs, and (c) "designed by and designed for the women" who are part of the service. (d) Gender-responsive interventions are guided by overarching goals to facilitate positive relationships in women's lives and to promote journeys toward critical autonomy. It should be noted that the use of CBT does not necessarily lead to a classification as either gender-responsive or gender-neutral. Rather, the context in which CBT is delivered is critical to the classification decision. Thus, multimodal programs which blend CBT with various therapies or skills training to address gender-specific needs were classified as gender-responsive, while interventions that were solely reliant on CBT to address one of the "Central 8" would be classified as gender-neutral.
4. **Environment (context)**—Gender-responsive intervention should be conducted (a) in a physically and psychologically, safe space; (b) trauma-informed practices and trained staff able to identify and understand the consequences of trauma. Criteria from both 4a and 4b must be met to be classified as gender-responsive. Notably, seven of the original studies included in the work by Gobeil et al. (2016) were reclassified under this framework. These are identified in Table 1 with a †.

Casework (Yes/No)

A binary variable, casework was an intervention involving a caseworker who assessed service user needs (either criminogenic or gendered), developed support plans, and signposted an individual to services and programs.

Location (Diversion, Institution, Community, Both)

A diversion intervention is usually a mandated program that diverts women from police custody or via specialist courts (Mental Health, Drug and Alcohol Courts) at the presentence stage into community services. Institution interventions take place in either a prison or secure residential units. Community interventions occur outside custodial settings and tend to follow a prison sentence but are not part of a diversion program. Interventions classed as "both" began in an institutional setting, usually prison, and then continued post-release with services or supervision delivered in the community.

Therapeutic Community (Yes/No)

These are participative, group-based approaches that create communities of service users and professionals to support recovery from the effects of mental illness and substance abuse.

SMS (3–5)

Studies with a comparison group were rated 3+. Studies were classed as 3, if they included a comparison group (but it is unmatched and is not statistically controlled as part of the analysis). Studies were classed as a 4 where a comparison group was matched or statistically controlled for in the analysis. Studies were classed as a 5 if they were randomized control trials.

TABLE 1 Summary of Studies

ID	Author	Focus	Outcome	Sample size						
				SMS	I	C	Location	Casework	Gendered	Follow-up
1	Armstrong et al. (2007)	Substance use	Charge / arrest	3	608	924	Community	No	Gendered	—
2	Brennan et al. (2018)	Multiple needs	Charge / arrest	3	44	114	Diversion	Yes	Gendered	12
3	Campbell et al. (2015)	Multiple needs	Charge / arrest	4	33	17	Diversion	Yes	Neutral	Variable
4	Cann (2006)—ETS	Cognitive skills	New conviction	4	52	156	Institution	No	Neutral	12
4	Cann (2006)—R&R	Cognitive skills	New conviction	4	66	198	Institution	No	Neutral	12
5	Carlson (2018)	Parenting	Any return	3	142	30	Institution	No	Partial	18
6	Caudy et al. (2018)	Multiple needs	Charge / arrest	4	2,556	996	Community	Yes	Neutral	—
7	Chan et al. (2005)	Multiple needs	Other	4	57	35	Community	Yes	Gendered	12
8	Costopoulos et al. (2017)	Multiple needs	Charge / arrest	4	15	25	Community	No	Neutral	6
9	Derkzen et al. (2017)—AWOCP	Multiple needs	New conviction	3	196	38	Both	No	Gendered	10
9	Derkzen et al. (2017)—WOCP	Multiple needs	New conviction	3	565	32	Both	No	Gendered	10
10	Dowden & Blanchette (2002)	Substance use	Not stated	3	58	40	Institution	Yes	Neutral	Variable
11	Durrance & Abitt (2001)	Multiple needs	New Conviction	3	61	19	Community	Yes	Gendered	12
12	Duwe & Clarke (2015)—C1	Cognitive skills	New Conviction	4	215	215	Institution	No	Partial	Variable
12	Duwe & Clarke (2015)—C2	Cognitive skills	New Conviction	4	860	860	Institution	No	Partial	Variable
13	Farrell (2000)	Substance use	Charge / arrest	5	42	38	Institution	No	Neutral	18
14	Freudenberg et al. (1998)	Multiple needs	Charge / arrest	4	51	142	Community	Yes	Partial	10
15†	Gat (2000)	Parenting	Any return	3	375	405	Institution	No	Neutral	60
16	Gehring et al. (2010)	Cognitive skills	Any return	4	111	111	Community	No	Gendered	12
17	George et al. (2015)	Multiple needs	Charge / arrest	4	269	331	Diversion	Yes	Partial	18
18	Gordon (2010)	Multiple needs	Charge / arrest	3	92	575	Both	No	Neutral	—
19	Grella & Rodriguez (2011)	Substance use	Any return	3	827	355	Both	No	Partial	12
20	Guydish et al. (2011)	Substance use	Charge / arrest	5	92	91	Community	Yes	Neutral	12
21	Hall et al. (2004)	Substance use	New conviction	3	101	79	Institution	No	Partial	12
22†	Harrell et al. (2001)	Substance use	Any return	3	110	26	Diversion	Yes	Neutral	6
23	Jason et al. (2016)	Substance use	Charge / arrest	3	86	84	Community	No	Neutral	24
24	Johnson et al. (2011)	Substance use	Any return	5	39	38	Community	No	Neutral	9
25†	Jollife et al. (2011)	Multiple needs	New conviction	4	660	660	Community	Yes	Gendered	1

(continued)

TABLE 1. (Continued)

ID	Author	Focus	Outcome	Sample size						
				SMS	I	C	Location	Casework	Gendered	Follow-up
26	Kubiak et al. (2016)	Multiple needs	Any return	3	19	16	Institution	No	Gendered	12
27	Kubiak et al. (2010)	Parenting	New conviction	3	48	36	Community	No	Gendered	120
28	Liau et al. (2004)	Cognitive skills	Not stated	5	38	32	Community	No	Neutral	6
29	Lindquist et al. (2009)	Multiple needs	Any return	3	153	204	Both	Yes	Neutral	12
30	MacSwain et al. (2014)—MMT-C	Substance use	Any return	3	24	45	Both	No	Neutral	6
31	MacSwain et al. (2014)—MMT-I	Substance use	Any return	3	61	45	Prison	No	Neutral	6
32	Matheson et al. (2009)	Substance use	Any return	3	318	108	Both	Yes	Gendered	12
33	McDonald & Arlinghaus (2014)	Multiple needs	New conviction	3	57	51	Both	Yes	Gendered	42
34	Messina et al. (2006)	Substance use	Any return	4	171	145	Institution	No	Neutral	12
35	Messina et al. (2012)	Substance use	Any return	5	85	65	Diversion	No	Gendered	—
36	Messina et al. (2010)	Substance use	Any return	5	60	55	Institution	No	Gendered	Variable
37	Miller et al. (2016)	Substance use	New conviction	4	32	30	Both	Yes	Neutral	12
38*	Miller (2010)—ATP	Substance use	Other	3	68	121	Community	No	Neutral	9
39*	Miller (2010)—T.C.	Substance use	Other	3	54	121	Institution	No	Neutral	9
40*	Miller (2010)—TC + ACP	Substance use	Other	3	122	121	Both	No	Neutral	9
41	Millson et al. (2010)	Multiple needs	Charge / arrest	5	174	174	Community	Yes	Gendered	12
42	Mosher & Phillips (2006)	Substance use	Any return	4	279	279	Institution	No	Neutral	6
43	Myer & Buchholz (2018)	Substance use	Other	4	63	63	Diversion	Yes	Gendered	24
44	Needles et al. (2005)	Multiple needs	New conviction	4	247	264	Community	Yes	Neutral	12
45	Nyamathi et al. (2018)	Cognitive skills	Other	5	58	58	Community	No	Neutral	6
46	Palmer et al. (2015)	Cognitive skills	New conviction	3	281	520	Community	No	Neutral	12
47†	Pelissier et al. (2003)	Substance use	Charge / arrest	3	331	142	Institution	No	Neutral	36
48	Prendergast et al. (1996)	Substance use	New conviction	3	19	23	Institution	No	Neutral	12
49	Richmond (2014)	Employment	Any return	4	1,685	1,685	Institution	No	Neutral	Variable
50	Robbins et al. (2009)	Substance use	Charge / arrest	3	108	116	Community	No	Neutral	18
51	Robertson et al. (2020)	Multiple needs	Any return	4	505	508	Diversion	Yes	Neutral	12
52	Roe-Sepowitz et al. (2011)	Prostitution	Charge / arrest	3	85	121	Diversion	Yes	Partial	12

(continued)

TABLE 1. (Continued)

ID	Author	Focus	Outcome	Sample size						
				SMS	I	C	Location	Casework	Gendered	Follow-up
53	Sacks et al. (2008)	Multiple needs	Charge / arrest	4	163	151	Institution	No	Gendered	6
54†	Sacks et al. (2012)	Substance use	Any return	5	207	163	Community	Yes	Neutral	12
55	Schram & Morash. (2002)	Multiple needs	Any return	3	147	69	Institution	No	Partial	2
56	Scott & Dennis (2012)	Substance use	Charge / arrest	5	224	238	Community	No	Neutral	3
57	Scott et al. (2017)—No probation	Substance use	Other	5	1,196	1,320	Community	No	Gendered	36
57	Scott et al. (2017)—Probation	Substance use	Other	5	1,001	983	Community	No	Gendered	36
58	Stalans et al. (2008)	Mental health	Charge / arrest	3	113	92	Community	Yes	Gendered	12
59	Sutherland (2019)	Multiple needs	Not stated	4	171	1,204	Community	No	Neutral	12
60	Torre & Fine (2005)	Education	New conviction	3	274	2,031	Institution	No	Neutral	36
61	Turning (2004)	Education	Charge / arrest	3	31	69	—	No	Neutral	36
62	Van Voorhis (2004)	Cognitive skills	Any return	4	742	749	Community	No	Neutral	12
63†	Watson et al. (2010)	Substance use	New conviction	4	164	173	Institution	Yes	Partial	12
64	Wexler et al. (1990)—Counseling	Substance use	Charge / arrest	3	113	38	Community	No	Partial	12
64	Wexler et al. (1990)—T.C.	Substance use	Charge / arrest	3	247	38	Community	No	Partial	12
65	Zarling et al. (2022)	Multiple needs	Charge / arrest	3	63	49	Community	No	Gendered	12
66†	Zlotnick et al. (2009)	Multiple needs	Any return	5	23	21	Both	No	Partial	6

Note. This table shows all of the studies included in the meta-analysis and how they were classified. Papers with multiple interventions were combined into a single pairwise comparison where the interventions were similar enough to do so. These are identified where the I.D. is the same for multiple interventions (e.g., Wexler et al., 1990). Studies containing multiple interventions to one comparison group are marked with * (Miller, 2010), meaning that in the analysis, the sample size of the control group was divided by the number of comparison groups. Studies marked with † indicate that it was reclassified from the work by Gobeil et al. (2016) regarding gender-responsive categorization (see Supplemental Material S5 for the full table and Supplemental Material S6 for the full reference list, available in the online version of this article). Columns "I" and "C" refer to the Intervention and corresponding comparison group's sample sizes, respectively.

Focus (Substance Abuse, Employment, Parenting, Mental Health, Housing/Accommodation, Emotion/Physical Abuse/Trauma, Multiple Needs)

Interventions were classified according to their primary focus. To be classed as multiple needs, the study must explicitly focus simultaneously on several needs or co-treatment needs. Studies focusing on a specific need but incidentally dealing with others were not classed as multiple needs.

Recidivism Measurement (Conviction, Charge/Arrest, Any Return to Custody, Not Defined)

Different studies adopt different measures of recidivism. There may also be differences for the same terms, such as "return to custody," depending on the study model. For example, "return to custody" in a probation setting may involve a failed drug test, whereas "return to custody" could result from a new crime. Similarly, certain studies band together charges/arrests, which are therefore not easy to delineate analytically. Some studies reported follow-up times and multiple measures of recidivism; to understand the implications of these ambiguities further, they were all extracted.

DATA ANALYSIS

Data analysis was divided into two parts: data pooling and subgroup analysis. In terms of data pooling, each study's effect sizes were calculated and then pooled to observe the overall pattern of the amalgamated data. Following this, to ascertain the validity and reliability of those overall patterns, the dataset's heterogeneity levels were calculated, and tests were conducted for publication and small-study biases in the data. After pooling the data, the subsequent analysis stage involved subgroup and meta-regression analysis.

Pooling Effect Sizes

Pooling effect sizes refer to statistically combining the effect sizes from multiple (similar) studies into an overall estimate of intervention effectiveness. The data analysis was conducted in R using the {*tidyverse*}, {*meta*}, and {*metafor*} packages. Once the data were extracted from all the papers into a complete dataset, the first stage of the analysis was to calculate odds ratios (ORs) using the raw counts for the sample size for experimental and control groups, and the recidivism counts per group. The {*esc*} package calculated the OR, variance, and standard error (all returned on the log scale). The package also returned the confidence intervals and effect size, which were then used to pool the effect size and heterogeneity using the {*metafor*} package.

Some studies reported results based on different follow-up periods and different measures of recidivism. The pooled effect sizes were chosen based on similarity to ensure the studies were comparable. The outcome measures pooled for each study were those with a follow-up period closest to 12 months based on new convictions.

Publication Bias and Small-Study Bias

A visual inspection of the funnel plot can only identify small-study effects and not tell us if publication bias exists. Egger's regression was used to test for publication bias using the {*dmetar*} package, and the small-study bias was also confirmed with a Likelihood Ratio Test (Egger et al., 1997). The funnel plot can be viewed in the Supplemental Material S2

(available in the online version of this article). This may also indicate the presence of small-study effects, but this may have been affected by significant heterogeneity.

Measuring Heterogeneity of the Pooled Data

A major challenge in meta-analyses is selecting sufficiently similar studies for data combination, but variability among pooled data remains a concern. High heterogeneity in the results suggests lower generalizability and reliability. Larger studies tend to influence the pooled average more due to their size, and the weighted average of study-level effect sizes is calculated to determine the strength of the experimental effect. Effect sizes were pooled by using the Mantel–Haenszel method and the Sidik–Jonkman estimator (Sidik & Jonkman, 2007) using the inverse variance method, the Paule–Mandel estimator for τ^2, and the Q-profile method for the confidence interval of τ and τ^2. A continuity correction of 0.5 was applied to studies with zero-cell frequencies to adjust for bias.

Both fixed- and random-effects models have been conducted and reported in the results section. According to the fixed-effect model, each effect size originates from a single, homogeneous population, and it asserts that the actual effect size is the same across all research. Finding a collection of entirely homogeneous research is extremely rare, which remains true even when we adhere to best practices and use our PICO to narrow the area of our study as much as feasible with studies as similar as possible. Therefore, the random-effects models are generally chosen by convention. As such, conclusions and interpretations were drawn from the random-effects models.

Outliers and Influential Studies

Like any averaging, outliers can bias results. Extremely influential studies (i.e., with a large sample size or an "abnormally" strong result) can skew the overall average picture of the research. Outliers are data points that contribute significantly to effect size heterogeneity, generally 50% of the *Q* statistic or greater (Hanson & Bussière, 1998). A leave-one-out analysis was conducted in which the pooled effect was calculated, excluding one study at a time. As it is common for studies to produce exaggerated effect sizes, this can be a valuable way to investigate each study's influence on the overall effect size.

Other Pooling Issues

Ideally, meta-analyses should only contain one data point from each paper. However, the Cochrane Handbook (Deeks et al., 2023) recommends several methods to deal with the inclusion of multiple papers from the same study. Some papers contained multiple studies with different control groups, others with multiple studies with the same control group, and some studies with multiple interventions but no comparison group. Each of these introduces different statistical considerations for the meta-analysis.

Studies that were similar enough in terms of categorization could be collapsed into one pairwise comparison. In the studies that involved multiple interventions with one comparison group, the sample size of that control group was divided by the number of comparison groups to avoid overinflation due to double counting (Deeks et al., 2023).

Several studies included in this meta-analysis lacked a "treatment as usual" group for comparison and instead compared their intervention against another intervention. Unlike the

work by Gobeil et al. (2016), this meta-analysis does not explicitly reference a "treatment as usual" or "control" group but rather a "comparison" group. Although this is a potential source of bias, it is largely mitigated by assigning SMS scores, which allows the effects of sub-optimal comparison groups to be partialled out as part of the meta-regression.

Subgroup Analysis

A subgroup analysis involves subsetting the data by subgroup and pooling the effect sizes for each subgroup using the abovementioned method, meaning that each group's effect size will be calculated, which could then be compared with the other groups. After the overall pooling, a subgroup analysis was undertaken for each of the subgroups, for example, gendered (gendered vs. gender-neutral / partially gendered), the intervention focus (substance use, multiple needs, cognitive skills, parenting, employment, prostitution, mental health, and education), therapeutic versus non-therapeutic, pathway (diversion vs. post-custodial), intervention location (community vs. institution), intervention length, individual vs. group intervention, and age of the sample.

Meta-Regression

A random-effects meta-regression (similar to a standard regression) was tentatively undertaken to confirm subgroup analysis while controlling for other potential confounds. Key variables were decided in the protocol to ensure that it is not overfitted and avoid spurious significances. It was initially intended that the meta-regression analysis incorporated outcome variables (gendered/gender-neutral/gender partial groups, intervention pathways, and therapeutic interventions) and covariates (focus, follow-up duration, intervention length, recidivism measures, and study quality). However, due to the limited sample size, it was not feasible to include all covariates simultaneously. Multiple meta-regressions were performed, and the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to compare the models and select the ones with the optimal variables that best fit the data. This approach aimed to achieve a robust model fit with fewer predictors.

RESULTS

The "Results" section follows the same structure outlined in the methods, which involve exploring the overall trends and reliability of these trends, then undertaking a subgroup analysis, followed by a meta-regression. Each of the subgroup analyses is discussed in corresponding subsections.

OVERALL POOLED EFFECT SIZE

There are 108 effect sizes taken from 71 interventions extracted from 64 papers (Table 1). These 71 interventions were collapsed down where necessary to 66 effect sizes (c.f. "other pooling issues" section) and pooled into a combined sample size of $N = 18{,}407$ women in the Experimental Groups and $N = 19{,}170$ women in the Control Groups. These effect sizes originated from some studies reporting multiple statistics (e.g., arrests/return to jail at 6 months, 12 months). Of the 71 interventions, there were 37 gender-neutral, 13 gender-partial, and 21 gendered interventions. Of the gendered and partially gendered studies, 5 used a

TABLE 2 Subgroup Analysis

		Sample size		Fixed effects		Random effects				
Group/subgroup	*k*	*E*	*C*	*OR*	CI	*OR*	CI	I^2	CI	Q^w/Q^b
All studies	71	18,407	19,170	0.76	[0.72, 0.8]	0.67	[0.57, 0.79]	0.87	[0.84, 0.89]	525.99
Gendered										
All	71	18,407	19,170	0.76	[0.72, 0.8]	0.67	[0.57, 0.79]	0.87	[0.84, 0.89]	525.99
Partial	13	3,244	2,472	0.57	[0.49, 0.65]	0.51	[0.39, 0.68]	0.49	[0.03, 0.73]	23.53*
Neutral	37	9,501	11,602	0.74	[0.69, 0.79]	0.79	[0.62, 1.01]	0.899	[0.87, 0.92]	357.75***
Gendered	21	5,662	5,096	0.89	[0.81, 0.97]	0.60	[0.44, 0.82]	0.832	[0.75, 0.89]	118.86***
Between										6.14*
Casework										
All	71	18,407	19,170	0.76	[0.72, 0.8]	0.67	[0.57, 0.79]	0.87	[0.84, 0.89]	525.99
No	48	12,298	14,748	0.85	[0.8, 0.91]	0.69	[0.55, 0.85]	0.854	[0.81, 0.89]	322.58***
Yes	23	6,109	4,422	0.61	[0.55, 0.67]	0.64	[0.49, 0.83]	0.862	[0.8, 0.9]	159.03***
Between										0.2
Focus										
All	63	15,654	14,701	0.79	[0.74, 0.83]	0.71	[0.6, 0.84]	0.87	[0.84, 0.89]	467.24
Substance use	31	6,910	6,107	0.82	[0.76, 0.89]	0.69	[0.55, 0.85]	0.799	[0.72, 0.85]	148.95***
Multiple needs	23	6,321	5,695	0.67	[0.61, 0.73]	0.66	[0.48, 0.92]	0.904	[0.87, 0.93]	229.74***
Cognitive skills	9	2,423	2,899	1.03	[0.89, 1.19]	0.89	[0.47, 1.68]	0.866	[0.76, 0.92]	59.51***
Between										0.86
Pathway										
All	69	18,315	19,056	0.76	[0.72, 0.8]	0.67	[0.57, 0.79]	0.87	[0.84, 0.89]	522.06
Community	29	9,435	9,320	0.88	[0.82, 0.94]	0.82	[0.62, 1.08]	0.921	[0.9, 0.94]	352.7***
Institution	21	5,277	6,911	0.57	[0.5, 0.63]	0.53	[0.39, 0.72]	0.792	[0.69, 0.86]	96.35***
Diversion	8	1,194	1,245	0.64	[0.53, 0.76]	0.52	[0.33, 0.84]	0.561	[0.03, 0.8]	15.93*
Both	11	2,409	1,580	0.76	[0.64, 0.89]	0.75	[0.54, 1.05]	0.546	[0.11, 0.77]	22.05*
Between										6.82

(continued)

TABLE 2. (Continued)

		Sample size		Fixed effects		Random effects				
Group/subgroup	*k*	*E*	*C*	*OR*	CI	*OR*	CI	I^2	CI	Q^w/Q^b
Therapeutic										
All	71	18,407	19,170	0.76	[0.72, 0.8]	0.67	[0.57, 0.79]	0.87	[0.84, 0.89]	525.99
Therapeutic	25	4,369	3,853	0.6	[0.53, 0.67]	0.56	[0.45, 0.7]	0.554	[0.3, 0.72]	53.82***
Not therapeutic	46	14,038	15,317	0.8	[0.76, 0.85]	0.73	[0.59, 0.91]	0.901	[0.88, 0.92]	452.91***
Between										2.93
Intervention length										
All	34	11,935	11,233	0.76	[0.72, 0.8]	0.67	[0.57, 0.8]	0.87	[0.84, 0.89]	523.33
0–6 months	27	11,502	10,729	0.85	[0.8, 0.91]	0.71	[0.59, 0.86]	0.761	[0.68, 0.82]	175.66***
>12 months	7	6,874	8,372	0.63	[0.58, 0.69]	0.63	[0.46, 0.86]	0.92	[0.89, 0.94]	323.37***
Between										0.51
Approach										
All	71	18,407	19,170	0.76	[0.72, 0.8]	0.67	[0.57, 0.79]	0.87	[0.84, 0.89]	525.99
Group	41	8,376	10,387	0.76	[0.7, 0.82]	0.65	[0.53, 0.81]	0.817	[0.76, 0.86]	218.24***
Unknown	13	5,082	2,659	0.57	[0.51, 0.64]	0.54	[0.33, 0.89]	0.933	[0.9, 0.95]	179.88***
Individual	17	4,949	6,124	0.92	[0.84, 1.01]	0.83	[0.61, 1.12]	0.798	[0.68, 0.87]	79.15***
Between										2.9
Female only										
All	71	18,407	19,170	0.76	[0.72, 0.8]	0.67	[0.57, 0.79]	0.87	[0.84, 0.89]	525.99
No	23	5,965	4,372	0.66	[0.6, 0.73]	0.7	[0.51, 0.97]	0.905	[0.87, 0.93]	230.98***
Yes	48	12,442	14,798	0.8	[0.75, 0.85]	0.65	[0.54, 0.8]	0.831	[0.78, 0.87]	278.77***
Between										0.16

Note. This table shows the results of the subgroup analysis, indicating the number of studies in each subgroup (*k*) and sample sizes for the Experimental and Control groups. It also shows ORs and CIs from both fixed-and random-effects models, and heterogeneity (I^2) and variability between the effect sizes of different studies from within (Q^w) and between (Q^b) subgroups.
$p < .05$*. $p < .01$**. $p < .001$***.

TABLE 3 Meta-Regressions—Gendered, Focus, Length, and Therapeutic Interventions

Model/Variable	*OR*	*SE*	*z*	*p*	*ci.lb*	*ci.ub*	
Regression 1							
Intercept	0.6	0.23	–2.26	.02	0.38	0.93	*
Gender-neutral (ref)							
Gendered	0.58	0.17	–3.18	<.01	0.41	0.81	**
Partial	0.7	0.19	–1.87	.06	0.48	1.02	
Community (ref)							
Both	1.24	0.25	0.88	.38	0.76	2.03	
Diversion	0.67	0.25	–1.55	.12	0.41	1.11	
Institution	0.66	0.18	–2.28	.02	0.46	0.94	*
Follow-up (months)	1.00	0.00	0.95	.34	1.00	1.01	
Any return (ref)							
Charge or arrest	1.27	0.2	1.19	.23	0.86	1.88	
New conviction	1.49	0.2	2.02	.04	1.01	2.19	*
Other	1.24	0.24	0.9	.37	0.77	1.99	
SMS = 3 (ref)							
SMS = 4	1.48	0.18	2.2	.03	1.04	2.1	*
SMS = 5	1.55	0.22	1.98	.05	1.01	2.39	*
Regression 2							
Intercept	0.93	0.25	–0.29	.77	0.57	1.51	
Community (ref)							
Multiple needs	0.61	0.22	–2.27	.02	0.4	0.94	*
Substance use	0.64	0.23	–1.95	.05	0.41	1	
< 6 months (ref)							
≤12 months	0.92	0.16	–0.47	.64	0.67	1.28	
Not therapeutic (ref)							
Therapeutic	0.76	0.14	–1.94	.05	0.57	1	*
Follow-up (months)	1	0.01	0.37	.71	0.99	1.02	
Any return (ref)							
Charge or arrest	1.16	0.18	0.83	.41	0.82	1.65	
New conviction	1.63	0.2	2.51	.01	1.11	2.39	*
Other	5.21	0.41	4.06	<.01	2.35	11.58	**
SMS = 3 (ref)							
SMS = 4	1.04	0.15	0.24	.81	0.77	1.4	
SMS = 5	1.38	0.19	1.64	.1	0.94	2.01	

Note. This table shows the meta-regression results for two regression models. Regression 1 includes gender and location, with covariates, such as intervention location, length of follow-up, measure of recidivism, and quality of study (SMS). Regression 2 includes the variables focus, length, and therapeutic nature of the studies with covariates, such as intervention location, length of follow-up, measure of recidivism, and quality of study (SMS). The OR is the ratio of odds between each subgroup and the reference group, and it was calculated using the exponent of log(OR).
$p < .05$*. $p < .01$**.

gendered assessment, 32 dealt with gendered pathways and had gendered content, and 20 had gendered context.

The random-effects pooled OR of these 66 effect sizes, OR = .67, [.57; .79], $p < .001$; in other words, women who attended a tested intervention were 33% less likely to reoffend within the follow-up period when compared with the comparison groups (see Table 2 and accompanying forest plot in the Supplemental Material S1, available in the online version of

this article). However, there is considerable heterogeneity (I^2 = 87% [83.1%; 88.9%], τ^2 = .33 [.22; .53], Q = 525.97, df = 70, $p < .001$), which means that the results may not be consistent. However, the OR was consistent when using a restricted maximum likelihood estimator; the results were T.E. = .67, τ^2 = .36, which did not differ from the original estimates. No individual studies contributed more than 50% to the overall Q-value = 525.97 (Visualized as a Baujat Plot, Supplemental Material S3, available in the online version of this article and leave one out analysis [Supplemental Material S4, available in the online version of this article]). There was no small-study/publication bias present in the pooled data, indicated by a largely symmetrical funnel plot (Supplemental Material S2), confirmed with Egger's test (intercept = −0.931 CI [−2.09; − .23], t = −1.57, p = .12), and by a Likelihood Ratio Test: $\chi^2(df = 1) = .003$, p = .95.

META-REGRESSIONS

Two meta-regression models fit the data best (with the lowest AIC and BIC) to explore as many variables as possible. Follow-up refers to the length of time after the interventions where the women were monitored for recidivism (months), and the outcome refers to the measure of recidivism (e.g., return to jail, arrests/convictions). Pathway refers to institutional, post-institutional, community-based, or diversion-based interventions.

SUBGROUP ANALYSES

Gender-Responsive Versus Gender-Neutral

Subgroup analysis shows that partially gendered interventions are most effective, followed by fully gendered compared with gender-neutral interventions (gendered OR = .60 [.44; .82], partial OR = .51 [.39; .68], gender-neutral OR =.79 [.62; 1.01]). However, when controlling for location, length of follow-up, measure of recidivism, and study quality (SMS), the first meta-regression (Table 3), gender-responsive interventions are most effective (reducing offending by 42%) followed by partially gendered interventions (reducing offending by 30%) when compared with gender-neutral interventions.

A general pattern is that the higher-quality studies yielded more conservative estimates, as seen in the first regression model, with SMS = 4 and SMS = 5 being up to 48% and 55% less effective than SMS = 3. To reiterate, SMS 4 and 5 are higher-quality studies (compared with SMS = 3) that generally ensure a randomized control or statistical control between treatment and control groups. Both meta-regressions also show that studies that use new convictions are, on average, 49% and 63% more conservative, respectively, in their estimations of recidivism compared with return to prison. However, gendered studies were consistently most effective at reducing recidivism compared with gender-neutral and gender-partial ones, even when accounting for study quality and measures of recidivism in the regression model.

The first meta-regression metrics (Table 3): τ^2 (estimated amount of residual heterogeneity) = .14 (SE = .05); τ (square root of estimated τ^2 value) = .38; I^2 (residual heterogeneity / unaccounted variability) = 64.66%; H^2 (unaccounted variability / sampling variability) = 2.83, R^2 (amount of heterogeneity accounted for) = 54.44%, Test for Residual Heterogeneity = QE (df = 46) = 158.10, $p < .001$; Test of Moderators = QM (df = 11) = 39.00, p = .002.

Intervention Location

The subgroup analysis suggests that diversion (OR = .52 [.33, .84]) and institutional (OR = .53 [.39, .72]) interventions show the highest effectiveness, followed by interventions starting in institutions and extending into the community (Both OR = .75 [.54, 1.05]), and finally, community-based interventions (OR = .82 [.62, 1.08]). When controlling for the other covariates in the meta-regression, institution-based interventions are significantly more effective than community-based interventions. However, they are similarly effective to diversion interventions (tending toward significance). Notably, the diversion group has only 8 interventions compared with 21 institutional-based interventions.

Group Versus Individual

Although not significant in the subgroup analysis, the data tend toward group-based interventions (OR = .65 [.53, .81]) being 18% more effective compared with individual-based interventions (.83 [.61, 1.12]), and the heterogeneity is high for both group and individual interventions (I^2 = 82% and 80%, respectively). There are more than twice the number of group-based interventions compared with individual-based interventions, and 17 studies do not specify group or individual interventions. In this unknown group, the OR is even lower, which means that while this finding may be important, this result should be interpreted cautiously.

Therapeutic Community

In the second meta-regression (Table 3), therapeutic interventions were significantly more effective (24% on average) than non-therapeutic communities when controlling for follow-up, intervention length, recidivism measure, and SMS. This pattern also persisted in the subgroup analysis, where the pooled therapeutic interventions also have much less heterogeneity (therapeutic I^2 = 55%, diversion I^2 = 90%). It should be noted that 7 out of the 10 therapeutic community interventions focused on substance abuse, meaning that there is considerable overlap between the two; however, when controlling for studies that focus on substance abuse, interventions adopting the use of therapeutic communities are significantly more effective than not. Due to the small sample size, it is difficult to reliably calculate heterogeneity.

Intervention Length

Intervention length was included in this study as a covariate. However, it was not included in the first meta-regression (Table 3) because the best-fit model did not contain this variable. Notably, 27 interventions were up to 6 months long, 7 were more than 12 months, and less than 5 were between 6 and 12 months long (and therefore not grouped). The rest did not explicitly mention the length of their interventions. Although not significant, the results tend toward showing that 12+ month interventions were 12% more effective compared with 0- to 6-months interventions (12+ months OR = .63, [.46; .86], 0–6 months OR = .71, [.59; .86]). The meta-regression also showed this to be non-significant, but it may lack statistical power to reach significance.

Intervention Focus

The focus of the intervention was included in this study as a covariate but not in the first meta-regression (Table 3) because the best-fit model did not contain this variable. However, a subgroup analysis shows that addressing multiple needs interventions (OR = .66 [.48, .92]) is most effective, reducing recidivism by approximately 34%, followed by substance abuse (OR = .69 [.55, .85]) and cognitive skills interventions (OR = .89 [.47, 1.68]). This pattern, again, was present in the second meta-regression (Table 3), and multiple needs interventions were 3% more effective than substance use and 39% more effective than cognitive skills interventions.

The metrics for the regression (Table 3): τ^2 (estimated amount of residual heterogeneity) = .10 (SE = .03); τ (square root of estimated τ^2 value) = .30; I^2 (residual heterogeneity / unaccounted variability) = 53.10%; H^2 (unaccounted variability / sampling variability) = 2.13, R^2 (amount of heterogeneity accounted for) = 67.49%, Test for Residual Heterogeneity = QE (df = 46) = 106.36, $p < .001$; Test of Moderators = QM (df = 10) = 42.93, p = .001.

DISCUSSION

In this section, we highlight key points from our analysis to understand "what works best" for CJI women in reducing recidivism. Our analysis has sought to identify the most effective approach (either gender-responsive or gender-neutral) and the different aspects of interventions that reduce the odds of reoffending. We are hindered to some degree by the lack of detail in the description of interventions, making it difficult to accurately identify the aspects of interventions that influence their effectiveness. Allied to this, studies tend not to report participant characteristics, so it is not possible to comment on "who" specific interventions might work for according to different demographic factors, needs/strengths profiles, and personal histories.

A GENDER-RESPONSIVE APPROACH WORKS BEST

Like the work by Gobeil et al. (2016), we demonstrate that gender-responsive interventions yield significantly reduced recidivism rates compared with the gender-neutral ones—approximately a 42% reduction when controlling for all other covariates. That said, unlike the work by Gobeil et al., our study showed a persistent significance, even when controlling for study quality, measures of recidivism (e.g., arrests, new convictions, and return to prison), intervention length, length of follow-up, and location.

We categorized gender-responsive studies according to four criteria. These criteria allowed us to explore the elements of the gender-responsive approach that might drive reductions in offending. It was not possible to explore subgroups of the individual gendered features (gendered pathways, intervention, context, and content) because in practice, these elements existed in a variety of different combinations, and ultimately, the sample size was not large enough to support such analyses. However, the results show the cumulative importance of these features; fully gender-responsive interventions were more effective than partially gendered ones, which were, in turn, more effective than gender-neutral ones. In other words, the closer the adherence of an intervention to the gender-responsive criteria discussed above, the more effective it is likely to be in reducing female offending.

The analysis also explored the format and intervention focus across both gender-responsive and gender-neutral approaches. Although not significant, group-based interventions can potentially be 18% more effective than interventions with individuals. Likewise, according to our analyses, therapeutic community rather than non-therapeutic community interventions demonstrate greater effectiveness in reducing recidivism for CJI women. These findings align with the principles of the gender-responsive approach, pointing to the importance of positive connections to others for recovery; to rebuild lives through peer relationships. Interventions focused on multiple needs were significantly (39%) more effective for women than cognitive skills-based (least effective) interventions. It should be noted that "cognitive skills" as an intervention practice was the only method to evidence no effect. This finding should be approached with some caution given that many of the interventions were not documented in sufficient detail to pinpoint the extent to which these interventions adhere strictly to the principles of CBT or blend other approaches into these programs. However, based on the information provided about these interventions, it could be speculated that CBT used in isolation may be less effective for CJI women and lends credence to the suggestion that multimodal approaches are more effective for this group.

INTERVENTION LOCATION

As with Gobeil et al.'s (2016) meta-analysis, institution-based interventions are more effective in reducing recidivism than those conducted in the community. This is the second meta-analysis focusing on women that has produced this finding, and it runs counter to the extant literature. Several hidden biases could contribute to this; one potential explanation for this pattern may lie in intervention dosage. While prison-based interventions are completed in a controlled environment, individuals in community-based programs remain at risk of reoffending during their treatment (c.f. Sutherland, 2019), potentially interrupting the full intervention dosage. The studies generally do not provide enough detail to control for this in this meta-analysis, however, dosage appears to be a critical factor; program completion is consistently associated with reduced recidivism rates in the literature that compares completers and non-completers.

Schemes that divert women from custody or via drug, mental health, or problem-solving courts that avoid prison, according to our analysis, are similarly effective in reducing offending to interventions that include prison. By definition, these are also community-based interventions but appear to be more effective than post-prison community-based interventions. Instinctively, this finding has an underlying logic. A key characteristic of diversion-based interventions was the structured pathways to multi-agency provision, resulting in women receiving support that addressed interrelated needs that brought them to the criminal justice system in the first place, while simultaneously avoiding the harms of imprisonment. Ultimately, the effectiveness of diversion interventions is likely to be determined by the quality and appropriateness of services that women are referred to.

INTERVENTION DOSAGE

It was not possible to explore dosage in detail due to omissions in the study reporting. It was, however, possible to examine the impact of intervention length, which still offered important insights. From the subgroup analysis, interventions more than 12 months in length were approximately 12% more effective than interventions of less than 6 months.

Thus, dosage is likely to be a significant factor, given the complexity of needs, such as trauma, mental health, and substance use faced by many CJI women, that often sit in the background behind "criminogenic needs" (Jason et al., 2016). Recovery and desistance take time and are not linear processes; they are often marked by setbacks that require patience and understanding—in fact, setbacks may be an important part of the journey (Barr, 2018). However, this does not necessarily undermine the assertion that appropriateness of intervention content rather than duration, might in some cases be a more significant factor in effectiveness (Messina & Esparza, 2022). Rather, it is perhaps more accurate to draw distinctions between shorter, targeted interventions that serve to stabilize substance use, mental health or offending behaviors, and longer interventions which aim to achieve sustained forms of recovery and significant changes to life circumstances. Ultimately, dosage will be determined by the end goal of the intervention.

METHODOLOGICAL ISSUES FOR FUTURE RESEARCH

Several methodological issues in the pooled studies. First, nuanced evaluation of intervention effectiveness is prevented by poor reporting of intervention dose, duration, and frequency (McGregor et al., 2016). Second, short follow-up periods of 6- to 12-months limit assessment of long-term effectiveness and insights into reoffending patterns (Lart et al., 2008). Third, as Sutherland (2019) suggests, matched groups are essential for correct interpretation; studies without a suitable control group can significantly alter the calculated efficacy of intervention. In the current research, higher-quality studies consistently produce more conservative estimates. Fourth, the choice of recidivism measures can significantly influence the estimates of effectiveness. Studies opting for a return to jail as their metric might appear less effective than those using new convictions. Fifth, recidivism presented as a single event is a blunt measure for a series of processes (Lart et al., 2008); binary reoffending rates fail to capture the complex and iterative pathways toward desistance that many women take (Barr, 2018). Broader measures of needs and well-being are necessary to understand the effects of interventions on reoffending (Sutherland, 2019). Finally, the included papers represent nearly 30 years of research, meaning that "sex" and "gender" have been used interchangeably. Although they claim to examine gender-responsiveness, their operational definitions and analyses are predominantly rooted in biological sex rather than gender. This meta-analysis has retained the definitions of "gender-responsive" consistent with the papers included; however, we recognize the need for future work to thoughtfully disentangle these constructs.

CONCLUSION

Our findings demonstrate clear evidence that gender-responsive interventions have lower recidivism rates than gender-neutral practices for CJI women. Our classification of gender-responsive interventions demonstrates that the greater the degree to which interventions are aligned to the components of a gender-responsive approach, the more effective the intervention is in reducing recidivism. Added to this, we identify forms and types of intervention that appear to be effective for CJI women regardless of whether they are gender-responsive, namely, the importance of a focus on multiple needs, structured forms of diversion from prison, therapeutic environments, and peer groups. Due to the paucity of reporting, it is

impossible to comment fully on which subgroups of CJI women are most likely to be positively affected by gender-responsive approaches. This must be the focus of future research studies and meta-analyses in this area.

ORCID ID

Richard Summers https://orcid.org/0009-0008-3869-0664
Simon Pemberton https://orcid.org/0000-0003-3485-7929
Joanna Long https://orcid.org/0009-0004-4372-8429

SUPPLEMENTAL MATERIAL

Supplemental Materials S1, S2, S3, S4, S5, and S6 are available in the online version of this article at http://journals.sagepub.com/home/cjb

REFERENCES

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

Barr, U. M. (2018). Gendered assisted desistance: A decade from Corston. *Safer Communities*, *17*(2), 81–93. https://doi.org/10.1108/SC-05-2017-0016

Belisle, L. A., Salisbury, E. J., & Keen, J. (2022). Did they move on? An outcome evaluation of the gender responsive program, girls. . . moving on. *Feminist Criminology*, *17*(2), 223–251. https://doi.org/10.1177/15570851211025401

Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of women offenders: An integrative perspective*. Wiley. https://doi.org/10.1002/9780470713013

Bloom, B., Owen, B. A., Covington, S., & Raeder, M. (2003). *Gender-responsive strategies: Research, practice, and guiding principles for women offenders*. U.S. Department of Justice, National Institute of Corrections. https://nicic.gov/gender-responsive-strategies-research-practice-and-guiding-principles-women-offenders

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/10.1177/0093854812456777

Brennan, T., & Jackson, E. (2022). Women's pathways: Replication and generalizability across state prison systems. *Criminal Justice and Behavior*, *49*(9), 1323–1341. https://doi.org/10.1177/00938548221102554

Covington, S. (1998). The relational theory of women's psychological development: Implications for the criminal justice system. In R. Zaplin (Ed.), *Female offenders: Critical perspectives and effective intervention* (pp. 113–131). Aspen Publishers.

Covington, S., & Surrey, J. (1997). The relational model of women's psychological development: Implications for substance abuse.In S. Wilsnack & R. Wilsnack (Eds.), *Gender and alcohol: Individual and social perspectives* (pp. 335–351). Rutgers University Press.

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*(1), 11–52.

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2023). Chapter 10 Analyzing data undertaking meta-analyses. 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 6.4*. Retrieved June 19, 2024, from https://training.cochrane.org/handbook/current/chapter-10

Eddy, J. M., Martinez, C. R., & Burraston, B. (2013). A randomized controlled trial of a parent management training program for incarcerated parents: Proximal impacts. *Monographs of the Society for Research in Child Development*, *78*(3), 75–93. https://www.jstor.org/stable/43773256

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Gehring, K., Van Voorhis, P., & Bell, V. (2010). "What works" for female probationers? An evaluation of the Moving On program. *Women, Girls, and Criminal Justice*, *11*(1), 6–10.

Gobeil, R., Blanchette, K., & Stewart, L. (2016). A meta-analytic review of correctional interventions for women offenders: Gender neutral versus gender-informed approaches. *Criminal Justice and Behavior*, *43*(3), 301–322. https://doi.org/10.1177/0093854815621100

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*(2), 348–362. https://doi.org/10.1037/0022-006X.66.2.348

Herman, J. L. (1992). *Trauma and recovery: The aftermath of violence—From domestic abuse to political terror*. Basic Books.

Jason, L. A., Salina, D., & Ram, D. (2016). Oxford recovery housing: Length of stay correlated with improved outcomes for women previously involved with the criminal justice system. *Substance Abuse*, *37*(1), 248–254.

Lart, R., Pantazis, C., Pemberton, S., Turner, W., & Almeida, C. (2008). Interventions aimed at reducing reoffending in female offenders: A rapid evidence assessment (REA). *Ministry of Justice Research Series*, *8*(08), 2008.

Mattos, C. T., & Ruellas, A. C. (2015). Systematic review and meta-analysis: What are the implications in the clinical practice? *Dental Press Journal of Orthodontics*, *20*(1), 17–19. https://doi.org/10.1590/2176-9451.20.1.017-019.ebo

McGregor, B., Brown, E., Yan, F., Mitchell, C., Robinson, C., DeGroot, J., & Braithwaite, R. (2016). Program success of mental health clients in day reporting centers. *Journal of Health Care for the Poor and Underserved*, *27*(2), 194–213. https://doi.org/10.1353/hpu.2016.0051

Messina, N. P., Calhoun, S., & Warda, U. (2012). Gender responsive drug court treatment: A randomized controlled trial. *Criminal Justice and Behavior*, *39*(12), 1539–1558. https://doi.org/10.1177/0093854812453913

Messina, N. P., & Esparza, P. (2022). Poking the bear: The inapplicability of the RNR principles for justice-involved women. *Journal of Substance Abuse Treatment*, *140*, 108798. https://doi.org/10.1016/j.jsat.2022.108798

Messina, N. P., Grella, C. E., Cartier, J., & Torres, S. (2010). A randomized experimental study of gender responsive substance abuse treatment for women in prison. *Journal of Substance Abuse Treatment*, *38*(2), 97–107. https://doi.org/10.1016/j.jsat.2009.09.004

Miller, W., & Rollnick, S. (2012). *Motivational interviewing: Helping people change* (3rd ed.). The Guilford Press.

Millson, R. D., & Van Dieten, M. (2010). *Women offender case management model: Outcome evaluation* (NIC Accession Number 018017). National Institute of Corrections.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS MEDICINE*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morash, M., & Kashy, D. A. (2022). The relevance of women's economic marginalization to recidivism. *Criminal Justice and Behavior*, *49*(3), 330–349. https://doi.org/10.1177/00938548211037585

Morash, M., Kashy, D. A., Smith, S. W., & Cobbina, J. E. (2016). The connection of probation/parole officer actions to women offenders' recidivism. *Criminal Justice and Behavior*, *43*(4), 506–524. https://doi.org/10.1177/0093854815626490

Myer, A. J., & Buchholz, M. M. (2018). Examining the impact of a gender-specific drug court on recidivism. *Journal of Crime and Justice*, *41*(2), 206–221. https://doi.org/10.1080/0735648X.2017.1282374

Richie, B. E. (2001). Challenges incarcerated women face as they return to their communities: Findings from life history interviews. *Crime & Delinquency*, *47*(3), 368–389. https://doi.org/10.1177/0011128701047003005

Richie, B. E. (2018). *Compelled to crime: The gender entrapment of battered black women*. Routledge. https://doi.org/10.4324/9780429500480

Salisbury, E. J., Boppre, B., & Kelly, B. (2016). Gender responsive risk and need assessment: Implications for the treatment of justice-involved women.In F. Taxman (Ed.), *Handbook on risk and need assessment* (pp. 236–259). Routledge. https://doi.org/10.4324/9781315682327

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. D. (1998). *Preventing crime: What works, what doesn't, what's promising*. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Sidik, K., & Jonkman, J. N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine*, *26*(9), 1964–1981. https://doi.org/10.1002/sim.2688

Skeem, J., Eno Louden, J., Manchak, S., Vidal, S., & Haddad, E. (2009). Social networks and social control of probationers with co-occurring mental and substance abuse problems. *Law and Human Behavior*, *33*, 122–135. https://doi.org/10.1007/s10979-008-9140-1

Sutherland, A. (2019). A propensity score analysis of a community resettlement programme for women prisoners. *Criminology & Criminal Justice*, *19*(1), 115–132. https://doi.org/10.1177/1748895817741513

Van Voorhis, P. (2012). On behalf of women offenders: Women's place in the science of evidence-based practice. *Criminology & Public Policy*, *11*(1), 111–145. https://doi.org/10.1111/j.1745-9133.2012.00793.x

Van Voorhis, P. (2022). Defining and evaluating gender-responsive treatment. In S. L. Brown & L. Gelsthorpe (Eds.), *The Wiley handbook on what works with girls and women in conflict with the law: A critical review of theory, practice, and policy* (pp. 137–156). Wiley. https://doi.org/10.1002/9781119874065.ch11

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

Wallace, B. C., Small, K., Brodley, C. E., Lau, J., & Trikalinos, T. A. (2012, January 28). Deploying an interactive machine learning system in an evidence-based practice center: Abstrackr [Conference session]. ACM International Health Informatics Symposium (IHI), Miami, FL, United States, pp. 819–824. https://doi.org/10.1145/2110363.2110464

Zarling, A., Scheffert, R., & Russell, D. (2022). Predictors of Retention and Recidivism of Justice-Involved Women in a Community-Based Gender-Responsive CBT Program. *Criminal Justice and Behavior*, *49*(3), 291–310. https://doi.org/10.1177/00938548211040850

Richard Summers has a background in Cognitive Neuroscience and Population Health, and is currently a Research Fellow in the Women, Crime, and Justice Research Group at the University of Birmingham.

Simon Pemberton is currently a professor in Social Policy and Criminology at the University of Birmingham. He is the director of the Women, Crime, and Justice Research Group, and has published widely on aspects of social harm, crime and criminalization, and issues in relation to poverty and social exclusion.

Joanna Long has a background in Health Research and evaluation methodology. She is currently a senior research fellow in the Women, Crime, and Justice Research Group at the University of Birmingham.