Article

Gender and Community Supervision: Examining Differences in Violations, Sanctions, and Recidivism Outcomes

Christopher M. Campbell[1] iD **,**
Ryan M. Labrecque[2], Megan E. Mohler[3],
and Molly J. Christmann[4]

Abstract

While correctional scholarship emphasizes the importance of addressing dynamic risk factors in changing behavior, a debate continues on the role of gendered effects relating to such factors and supervision strategies. This study advances the debate and approach to violators by examining factors that reduce recidivism among technical violators. Using a large purposive sample of people on supervision, this study assesses differences between men and women in violating behavior, the sanctions used, and subsequent responses. To isolate effects of gender, we employ propensity score modeling coupled with a doubly robust design. Findings suggest that while some gendered differences exist, effect sizes are generally small and suggest responses to violators may be best focused on dynamic factors generally, rather than gender specific.

[1]Portland State University, OR, USA
[2]University of Central Florida, OR, USA
[3]Temple University, Philadelphia, PA, USA
[4]City of Portland, OR, USA

Corresponding Author:
Christopher M. Campbell, Criminology and Criminal Justice, College of Urban & Public Affairs, Portland State University, P.O. Box 751, 1600 SW 4th Avenue Suite 550, Portland, OR 97207-0751, USA.
Email: ccampbell@pdx.edu

Keywords

For the last several years, fiscal crises and greater expectations of evidence-based practices have nudged correctional agencies toward being less reliant on incarceration. In this new era of decarceration, a greater emphasis has been placed on supervising offenders in the community (Barker, 2011). Offenders who violate their conditions of supervision, however, have been shown to comprise more than a third of all prison admissions in some jurisdictions (e.g., Grattet et al., 2009). With such a large proportion of offenders failing on correctional supervision, agencies have struggled to identify the most effective methods for dealing with its technical violators.[1] Some common responses to rule violations have involved the use of graduated sanctions (Harrell & Roman, 2001; Taxman et al., 1999; Wodahl et al., 2013), swift-and-certain sanctions (Hamilton, Campbell, et al., 2016; Latimore et al., 2016; O'Connell et al., 2016), and rehabilitative programs (Bogue et al., 2004; Taxman & Belenko, 2012; Zhang et al., 2006).

While agencies search for the best approach for responding to technical violators (Grattet & Lin, 2016; Grattet et al., 2009; White et al., 2011; Wodahl et al.,2015) and reducing recidivism (Jolliffe & Hedderman, 2015; Kubrin & Stewart, 2006), a parallel issue questions whether there is a difference in how men and women offenders respond to correctional efforts (e.g., punishment and rehabilitation). Some scholars contend that the correctional system should provide a tailored approach to men and women offenders which addresses their distinct socialization experiences and pathways into crime (i.e., gender-responsive; e.g., Van Voorhis et al., 2010). An assumption of this position is that men and women respond to correctional interventions (which are traditionally created by and for men) differently because the two groups have different criminogenic needs and responsivity considerations based on their unique pathways. Other scholars, however, maintain that criminogenic needs are universal (i.e., gender-neutral) and insist the best strategy for eliciting behavioral conformity is adherence to the principles of effective intervention (Bonta & Andrews, 2017).

These competing perspectives have largely not been examined outside of risk prediction (e.g., Smith et al., 2009) and the gender-specific pathways literature (e.g., Brennan et al., 2012). In the context of community supervision, very few studies to date have investigated the moderating influence of gender on technical violations, sanctions, or recidivism. Such investigations are particularly scarce where the context of graduated sanctions and the principles of effective intervention are relied on as a response to violating

behavior. Considering these two approaches are tied to empirical research and are increasingly common in supervision practices, research must account for these considerations in examining baseline assumptions of gender differences. This gap in knowledge is important from both a theoretical and practical standpoint. If gender differences do not exist, then correctional agencies should treat the technical violations of men and women similarly. However, if such differences do exist, then authorities should consider implementing gender-specific supervision, sanctions, and treatment strategies in order to achieve the greatest reductions in recidivism. The current study weighs in on this debate by evaluating whether men and women parole violators in Washington State have different violating behavior, receive different sanctions for such behavior, and respond differently to sanctions (i.e., recidivate via new crime).

Literature Review

Responding to Violations

There has long been a struggle to understand how best to reduce the criminal behavior of offenders on community supervision. As many states continue to observe high rates of recidivism (Durose et al., 2014) and violation-related returns to incarceration (Grattet et al., 2009), there is a need for more effective offender management strategies. Recommendations have taken many forms, including applying various types of sanctions on violators, implementing the principles of effective intervention, and providing gender-responsive treatment services.

Graduated sanctions. In many jurisdictions, probation and parole officers respond to technical violations with graduate sanctions. Possible sanctions exist on a continuum and can be community-based or involve incarceration (Taxman et al., 1999).[2] For example, sanctions can take the form of increased control and surveillance such as using electronic monitoring devices. The research on electronic monitoring, however, has produced mixed results, particularly for parolees. Some studies have reported this practice has no effect on the likelihood of being recommitted to prison (Finn & Muirhead-Steves, 2002), while others have suggested it significantly reduces the likelihood of supervision failure, especially among certain offender types (e.g., violent offenders more than drug or property offenders; Padgett, Bales, & Blomberg, 2006). Another approach involves increased officer-offender contacts or participation in an intensive supervision program, for which studies have also produced varied results (Erwin, 1986; Petersilia & Turner,

1993). While some research has suggested offenders assigned to intensive supervision recidivated more often than those on traditional supervision (Gendreau et al., 2000), other studies have found that coupling intensive supervision with rehabilitation programming resulted in improved behavior (Lowenkamp et al., 2010).

The most severe type of sanction employed is confinement in jail or prison. Increasingly, sanctions are being justified on the philosophy of deterrence. This ideology is based on the idea that offenders will disengage from criminal activity when the punishment for such behavior is perceived as being swift, certain, and proportionally severe (Kleiman et al., 2014). Despite its promise, however, the use of swift-and-certain sanctions has produced mixed results with some studies reporting it reduced recidivism (Hamilton et al., 2016; Hawken et al., 2016), and others suggesting it produced null or iatrogenic effects (Latimore et al., 2016; O'Connell et al., 2016). Moreover, research on the use of jail and prison as a sanction has also shown varied results (Bucklen, 2014; Campbell, 2015; Drake & Aos, 2012; Wodahl et al., 2015). The use of incarceration sanctions rests on the assumption that most offenders are motivated to be more conforming, and wish to adhere to their conditions of supervision. While this philosophical approach to addressing violative behavior has struggled to find consensus among scholars, other approaches (e.g., rehabilitative treatment) have found more empirical and popular support in influencing behavioral change of offending populations.

Principles of effective intervention. Recent decades of reform have helped increase the implementation of the principles of effective intervention within many community supervision settings (Bogue et al., 2004; Bonta & Andrews, 2017; Gendreau et al., 2004). Among these guiding principles is the belief that actuarial risk assessments (i.e., data-driven, predictive tools) can be used to identify offender likelihood for recidivism, and further that the use of targeted correctional interventions can curtail one's risk (i.e., the risk-need-responsivity [RNR] approach; Labrecque et al., 2014). Under this framework, the most intensive interventions are reserved for the highest risk offenders with the expectation that this group will benefit the most. The types of interventions suggested will produce the greatest reductions in recidivism are those addressing the eight criminogenic needs (i.e., crime-producing factors): Education/employment, financial, family and marital status, living situation, anti-social friends/associates, alcohol/drug problems, mental health, and anti-social attitudes (Bonta & Andrews, 2017). Research has further suggested that the most effective types of interventions include cognitive behavioral therapy (Lipsey et al., 2007), drug and alcohol treatment (Prendergast et al., 2013; Taxman, 1999), halfway houses (Hamilton & Campbell, 2014; White

et al., 2011), day reporting centers (Ostermann, 2009; Spence & Haas, 2015), housing vouchers (Hamilton et al., 2015; Lutze et al., 2014), and specialty courts (Hamilton, 2011; Shaffer, 2011).

In spite of the strong empirical support for the principles of effective intervention generally (Bonta & Andrews, 2017), there has been surprisingly little research on its applicability with technical violators more specifically. While some graduated sanctions seemingly align with the tenants of these principles (e.g., connecting offenders with positive drug tests to appropriate substance abuse treatment services), others, including many deterrence-based strategies, are theoretically at odds with this correctional ideology. According to this perspective, sanctions matched to offender risk and needs are more effective than those that are applied similarly across all offenders. Furthermore, this position maintains that interventions emphasizing these principles are more effective in addressing violative behavior than those based solely on deterrence.

Gender-responsive perspective. In addition to the search for best practices regarding technical violators, a separate and simultaneous debate has emerged in corrections regarding gender-responsive approaches to reducing recidivism. On one side, some scholars contend the principles of effective intervention and the dynamic needs domains (e.g., education, substance abuse, family support) universally explain criminal behavior (Bonta & Andrews, 2017). On the other side, however, other scholars have posited a gender-responsive perspective. This position suggests men and women have fundamentally different life experiences and socialization processes (i.e., gendered pathways) that lead them into and out of criminal behavior (Bloom et al., 2004; Brennan et al., 2012; Wattanaporn & Holtfreter, 2014; Rettinger & Andrews, 2010; Salisbury & Van Voorhis, 2009; Yesberg et al., 2015). Due to these differing pathways, men and women offenders therefore require separate correctional approaches to best reduce recidivism.

This debate has important implications for the correctional system as a whole because in practice it relies heavily on predicting and mitigating recidivism through the use of offender risk-needs assessments. An important component for gauging responsivity and providing effective treatment, these risk assessments have received attention and scrutiny in the gender-neutral versus gender-responsive debate as it relates to risk level (e.g., high versus moderate risk) and dynamic risk factors or criminogenic needs (e.g., importance of antisocial family and childcare; Yesberg et al., 2015). The predictive validity of risk-needs assessments for both men and women have retained support across multiple studies, however, there is contradictory research pertaining to the validity of individual domains (Andrews et al., 2012; Ostermann &

Salerno, 2016; Rettinger & Andrews, 2010; Greiner, Law, & Brown, 2015; Smith et al., 2009; Van Voorhis et al., 2010). Some studies suggest that some domains important for predicting recidivism among men may not be as accurate when applied to women (Van Voorhis et al., 2010). In opposition, others contest that primary domains (Bonta & Andrews, 2017) are just as predictive for women as they are for men (Andrews et al., 2012; Smith et al., 2009; Rettinger & Andrews, 2010). Responding to this debate on risk, Smith et al. (2009) conducted a meta-analysis of 25 studies involving a popular risk/needs assessment (i.e., the Level of Service Inventory-Revised) and 14,737 women offenders. Smith et al. (2009) found that this instrument predicted recidivism well among both men and women.

In practice, the debate and its associated assumptions over gender responsivity/neutrality has important implications for the effective delivery of services and sanctioning for people on community supervision (Skeem et al., 2016; Yesberg et al., 2015). According to the gendered-responsive perspective, the overwhelming majority of correctional practices are not effective for women because they were designed for men and have been applied to women as an afterthought (Bloom et al., 2004; Salisbury et al., 2009; van Wormer & Kaplan, 2006; Yesberg et al., 2015). This gendered position suggests that the best way to improve outcomes for women on supervision is to incorporate women-specific interventions, such as trauma informed services, familial support, childcare, and more mental health treatment. While some research reports that gender-specific interventions are more effective with women offenders (Gobeil et al., 2016), other research suggests that more traditional treatment services (e.g., CBT programs) produce similar impacts among men and women (Andrews & Dowden, 1999; Dowden & Andrews, 2000; Lowenkamp & Latessa, 2005). Most recently, research has revealed that women may respond differently to treatment and sanctions depending on their risk level (Morash et al., 2019). In a study of 385 women on supervision, Morash et al. (2019) identified that low-risk women recidivated less within 24 months when provided a sanction following non-drug violations, but recidivated more when assigned a condition of treatment. The opposite effects were observed for high-risk women – treatment was associated with reduced recidivism, while sanctions were not.

Current Study

In spite of the advancements from these three correctional subfields, key gaps remain in order to develop a cumulative, or consensus, knowledge base for best practices in addressing violators and providing intervention efforts in supervision. While some research posits that predictions of risk may be

independent from measures of responsivity to certain treatment delivery, and that the nature of predictive measures can capture effects of gender, it does so without explicit parameters (Rettinger & Andrews, 2010). It remains to be seen if certain strategies are more effective than others, particularly depending on the offender's gender. In addition, gender-specific research on treatment efforts for women suggests women may benefit from interventions and treatments based on their specific, often gendered, needs (Gobeil et al., 2016; Messina et al., 2012; Saxena et al., 2014; van Wormer & Kaplan, 2006), however, much of the existing evidence focuses only on single-gender samples, or they lack critical comparison between the genders. If women indeed respond differently to treatment interventions, logically similar effects should be detectable in responses to violation behavior.[3] Little is currently known about gender differences when it comes to sanctioning violators. Moreover, extant research on sanction effectiveness acknowledges individualized sanctions may be important, but most evaluations tend to analyze largely male samples that are void of gender-responsiveness. Thus, there is a clear need to establish relationships between gender differences and sanction effectiveness for technical violations, especially as it relates to graduated sanctions that include treatment components. The current study aims to address these gaps by investigating the effect of violation sanctions on recidivism, and the differences between men and women. It does so by addressing three research questions: To what degree is there a difference between men and women offenders in (1) violating behavior, (2) the sanctions given for violations, and (3) how they respond to such sanctions?

Method

To address these questions, this study analyzed administrative data provided by the Washington State's Department of Corrections (WADOC). In order to isolate any potential gendered effects of sanctions, a quasi-experimental design using propensity score modeling (PSM) was employed to balance the groups on theoretically important covariates for comparison. Specifically, men offenders were matched to women offenders by way of static and dynamic risk factors, first violation type, and first sanction type. Once balanced on these covariates, binary logistic regression and survival analysis were used to assess if gender is an important predictor of recidivism and time-to-recidivistic event. This design provided the opportunity to assess the existence of gender differences in systemic responses (i.e., sanctions) to violations. If gender is not related to subsequent violations, sanctioning, or recidivism, then it should not be predictive of any outcomes. However, if there are differences related to gender, then such differences should be apparent in their predictive nature after the match.

Eligibility

This study involved a purposive sample of parolees[4] released from the WADOC between 2008 and 2010 with a minimum follow-up period of 2 years. The sampling time frame was chosen for two reasons. First, the WADOC did not collect dynamic risk (i.e., criminogenic needs) information until 2008. Given the literature debate on gendered criminogenic needs, this study focused its sample on cases released after 2008. Second, the 2010 year was chosen as a final year of release to ensure sufficient follow-up time before a noted policy change in Washington.[5] These criteria resulted in an eligible pool of 54,540 cases. With our focus on individual differences in violations and sanctions, we narrowed the eligibility criteria to include only individuals who experienced at least one formal violation-sanction event. That is, the parolee must have committed at least one violation that was documented by the community corrections officer and was subsequently sanctioned (including a discretionary but documented "no sanction" and/or verbal reprimand). This criterion reduced the eligible sample to 28,509 cases. Lastly, due to various administrative reasons in WADOC data collection, a number of cases had missing values in the criminogenic needs or static risk information. After noting no discernable pattern in the missing values that would indicate a concern (i.e., the data were missing completely at random), these cases were removed from the eligible pool, producing a pre-match sample of 22,106 cases, including 18,886 men and 3,220 women.

During discussions about the study, the WADOC administrative and research staff expressed a concern that focusing on the first sanction event alone would be too narrow (Harrison, 2017). The WADOC officials expressed a belief that parolees do not necessarily improve their behavior after only one sanction, and that it may take two or more sanctions to reach this desired result. Upon inspection, the concern about multiple sanctions per offender was corroborated in the data. The average number of sanctions experienced by both men and women in this sample was four (standard deviation of about three, see Table 1). In order to account for this potential, this study's analyses focus on both the first and second violation-sanction events experienced by each person, including the highest or most severe violation listed for the associated sanction.

Measures

Men and women were compared on 337 covariates prior to the propensity score analysis. The measures came from three categories: Risk-need items, technical violations, and sanctions.

Table 1. Violation and sanction types by gender.

Violation type	%Men	%Women	χ^2	p	d	95% CI
Total event observations (n)	142,027	25,442				
Low-level violations						
General low-level	12.1	11.1	18.9	<.001	0.021	[0.012, 0.031]
Employment related	0.2	0.1	13.2	<.001	0.018	[0.008, 0.027]
Fine, restitution, community service	5.1	5.0	0.9	.349	0.005	[−0.005, 0.014]
Geographic or electronic monitor	8.7	8.7	0.1	.967	0.001	[−0.009, 0.010]
Failure to attend treatment	2.7	2.9	3.4	.067	0.009	[−0.001, 0.019]
Drug related	36.6	41.8	247.8	<.001	0.077	[0.067, 0.087]
High-level violations						
General high-level	1.5	0.8	65.1	<.001	0.030	[0.030, 0.049]
Sex related	1.1	0.1	249.4	<.001	0.077	[0.068, 0.087]
Weapons use/possession	0.2	0.1	12.1	<.001	0.017	[0.007, 0.027]
Prohibited contact	2.1	1.5	45.1	<.001	0.033	[0.023, 0.042]
Absconding	19.0	19.8	8.1	.005	0.014	[0.004, 0.024]
Rearrest						
Misdemeanor	6.4	5.2	51.1	<.001	0.035	[0.025, 0.045]
Felony	4.2	2.9	89.1	<.001	0.046	[0.037, 0.056]

	Mean (SD)		t	p	d	95%CI
Violations per person	7.7 (7.0)	8.0 (7.3)	5.9	<.001	0.040	[0.027, 0.054]

Table 1. (continued)

Response type	%Men	%Women	χ^2	p	d	95% CI
Total event observations (n)	71,512	12,303				
No-sanction/reprimand	26.9	29.5	36.8	<.001	0.042	[0.028, 0.056]
ERP	5.2	5.8	6.7	.010	0.018	[0.004, 0.031]
Reparation/fees/fine	1.2	0.8	15.0	<.001	0.027	[0.013, 0.040]
Opt. treatment or targeted intervention	3.4	3.7	2.6	.109	0.011	[−0.002, 0.025]
Inpatient treatment	0.8	1.3	28.7	<.001	0.037	[0.024, 0.051]
Work release	0.3	0.8	71.1	<.001	0.058	[0.045, 0.072]
Jail	60.6	56.6	68.6	<.001	0.057	[0.044, 0.071]
Prison	1.6	1.5	0.6	.451	0.005	[−0.008, 0.019]
	Mean (SD)		t	p	d	95% CI
Sanctions experienced	3.8 (3.1)	3.9 (3.4)	2.7	.007	0.026	[0.007, 0.046]
Violations stacked per sanction	4.4 (4.1)	4.3 (3.8)	0.3	.749	0.003	[−0.016, 0.022]

Note. All p values are for chi-square (χ^2) or t values, and confidence intervals (CI) are for Cohen's d (d).

Risk-needs items. This group of covariates involved individual items from Washington State's Static Risk Assessment (SRA) and Offender Needs Assessment (ONA).[6] The SRA primarily consists of criminal history information and was used to determine one's institutional classification level and supervision intensity. The ONA includes criminogenic need information on education, employment, friends/associates, residence, family, substance use/abuse, mental health, aggression, attitudes/behaviors, and coping. While the ONA was not a direct product of the principles of effective intervention (e.g., Bonta & Andrews, 2017) or the gender-responsive perspective (e.g., Van Voorhis et al., 2010), it collects rather nuanced information that would be found in both camps. For example, in the family domain of the ONA, while there are questions on the criminality of one's spouse or partner (as would be found in the literature on principles of effective intervention), there are also questions on the presence of minor children who reside with the incarcerated parent (noted as important in the gender-responsive literature). The items from the SRA and ONA were dichotomized to capture nuanced differences between men and women within each domain.[7] Due to the large number of covariates available, only the key demographics and measures that remained unbalanced after the match are discussed in later sections regarding the bivariate analysis of pre- and post-match differences.[8]

Technical violations. This group of covariates involved information about violations and associated sanctions. During the observation period, types of technical violations in Washington State ranged widely. Due to the scattered distribution across some less frequent violations, we examined this information in three state-defined groupings. The first involved low-level violations, which included general low-level, employment related, fines/restitution/community service, geographic or electronic monitoring related, failure to attend treatment, and drug related (e.g., positive urinary analysis) violations. The second was high-level violations[9] which included categories of general high-level, sex related, weapons use/possession, and contact with prohibited locations or people. Finally, absconding from supervision was kept as a distinct violation despite being categorized by the department as a high-level violation. This decision was largely due to the fact that absconders commonly (and in this sample) made up a large group of violators and have been shown to have some distinct differences (e.g., Hamilton & Campbell, 2013).

Sanctions. This group of covariates included information about the type and frequency of sanctions for technical violations. During the observation period, the WADOC used a graduated sanctioning grid to aide community corrections officers in determining an appropriate sanction. One to six

low-level violations could be met with the following graduated options: A discretionary "no sanction" or verbal reprimand, referral to some type of treatment or intervention targeting criminogenic needs,[10] condition enhancement which could include additional supervision contacts or check-ins, fines or reparation payments, partial confinement (e.g., work release or home detention), and jail. More than six low-level violations or any single high-level violation was subject to receiving up to 60 days in jail per violation-sanction event (totaling up to 300 days or more) and/or a full revocation by way of hearing (i.e., return to prison for the remainder of the sentence; Drake & Aos, 2012). These options were collapsed into five categories: (1) No-sanction or reprimand, (2) treatment or targeted intervention, (3) enhancement, reparation, or partial confinement (ERP), (4) jail, and (5) prison.[11] With jail and prison included in an analysis of recidivistic events, time-at-risk becomes a concern that must be accounted for as those who are incarcerated are not at-risk to violate or commit a new crime. Thus, a measure termed "exposure" was created to account for days spent incarcerated due to a violation sanction.

Outcome measures. Multiple outcomes were included to explore for potential differences in the impact of sanctions between men and women. The outcomes examined here involve four types, including subsequent technical violation, rearrest, reincarceration, and new conviction. Technical violations were collapsed into dichotomous categories of any low-level violation, high-level violation, and absconding. The rearrest measure counted any new arrest for a misdemeanor or felony while on supervision and was therefore given a technical violation for the standard condition "must obey all laws." Reincarceration included any return to jail or prison for any reason, including revocation. Jail and prison measures were analyzed separately because they differed substantially by the length of stay upon confinement. New convictions involved any new misdemeanor or felony conviction brought by the state's prosecution. Lastly, all events requiring custodial resources (e.g., rearrest, reincarceration in jail or prison, and new conviction) were combined into a measure of "any return to custody."[12]

Analytical Plan

In order to isolate the potential differences between men and women, we employed PSM followed by bivariate analyses. Such PSM approaches have been increasingly relied on in community correctional contexts (e.g., Vito et al., 2015). Generally, PSM is used to remove selection bias that exists between groups in the aim of simulating a randomized controlled trial or

experimental design. As a result, PSM provides a quasi-experimental approach with a compatible comparison group that allows the researcher to estimate an unbiased average effect of one group's condition in relation to the control group (Guo & Fraser, 2014). This is accomplished by matching cases on their propensity to be in one group over the other (i.e., their propensity score; Rosenbaum & Rubin, 1983). The propensity score provides a summary of how the covariates are related to the group of interest (Leite, 2017).

The use of PSM was employed in this study to balance and isolate the effects of gender (see Guo & Fraser, 2014; Lee, 2011). Recent scholarship has shown PSM to be an effective strategy for matching/balancing groups on biological traits and estimating treatment effects in criminal justice settings (e.g., matching to isolate the effects of race, see Franklin, 2015). We opted to use PSM in this way for three reasons. First, by matching men to women, this process highlights potential similarities and differences that may help contribute to the gender debate. Second, this strategy provides a more detailed and unbiased comparison between men and women, accounting for any covariate potential differences that may be inherent to a gender (e.g., criminal history and criminogenic needs). Lastly, conditioning the propensity score essentially creates a summary of so-called gender-neutral covariates. While it could be argued that fitting the appropriate model to capture socialization processes of gender would require a much fuller set of covariates, we argue that this is precisely addressing the noted gender debate. Any difficulties in matching or prediction would suggest that these measures are not gender-neutral. After matching on the covariates, the bivariate examination of outcomes then provides an unbiased estimate of differences in the effects of sanctions on men and women.

There were six metrics used to assess the quality of the match (see Campbell & Labrecque, 2018). Considering the goal of PSM is to mimic the effects of random assignment, the percent of covariates that possessed a statistically significant difference ($p < .05$) were assessed. A balanced sample should possess fewer than 5% of the observed covariates with statistically significant differences. As significance tests can be volatile due to sample size, another measure of balance used is the standardized percent bias (i.e., % bias; Rosenbaum & Rubin, 1985). The standardized percent bias provides a value indicating the degree to which proportions or means of each covariate are different between groups. By convention, it is expected that balanced covariates will possess a value of less than 10% absolute bias following the application of PSM, but traditionally 20% or less has been viewed as acceptable (Austin, 2011). In this study, we examined the overall mean and maximum standardized percent bias, as well as the percent of total covariates over 10% and 20% bias. Lastly, the area under the curve (AUC)

statistic was calculated with a receiver operating characteristic to test the accuracy of the estimated propensity scores to predict the case's gender based on the covariates used in the propensity score. This statistic has been used as a multivariate assessment of PSM elsewhere (e.g., Hamilton & Campbell, 2014; Hamilton et al., 2013). The value of an AUC statistic ranges between 0.5 and 1, where 1 is a perfect prediction of a true positive, and 0.5 is random chance (Cook, 2008). According to Rice and Harris (2005) an AUC value is strong if it is above 0.71, moderate if between 0.64 and 0.71, and weak if between 0.56 and 0.64. Ideally, the pre-match AUC should be an accurate predictor of receiving treatment (i.e., 0.7 or above to predict that the case is a woman), while post-match AUC values should be closer to 0.5 indicating that the groups are balanced. Bivariate post-match examinations relied on Cohen's *d* as the primary effect size, which can be interpreted following Cohen's (1988) guidelines.

Following a successful match, we split the post-match analysis into a second violation-sanction event (immediately following the first, which was used in the match) and any subsequent recidivistic event (i.e., anything immediately following the second experience). A series of logistic and Cox proportional hazard regressions were used to isolate the ability of gender to predict outcomes identified as remaining significantly different after the match. Additional regression models following PSM have been shown to be beneficial in isolating effects and importance of other covariates (Kurth et al., 2006), and is increasingly expected as a "doubly robust" approach (Stuart, 2010). For the current study, the use of a post-match, binary logistic and Cox regression allowed us to examine the importance of gender in predicting outcomes while controlling for the type of sanction given and any unbalanced covariates. As there are a number of measures used in this complex study, it is important to note that while PSM was used to balance out the risk and needs covariates between men and women, post-PSM regressions only incorporate those covariates that remain unbalanced. Moreover, testing the interaction of gender and sanction types within each model provided a more well-rounded examination of ways that gender might influence the outcome of supervision.

Results

Pre-Match Analyses

To address the first two research questions, the data was analyzed in a long format (i.e., one-to-many). This strategy provided a structure in the data emphasizing the event (violation or sanction) as the unit of analysis. In other

words, a person is listed for every violation he or she commits. This structure also afforded the opportunity to examine differences in the types of violations and sanctions by gender. This process exposed 167,469 violation observations, with approximately 85% committed by men. These violations resulted in 83,815 sanction events (including verbal reprimands), with a similar distribution of approximately 85% going to men. The bivariate analyses between the violations and sanctions are provided in Table 1.

Most violation types were statistically different between men and women ($p < .001$). Due to the large number of observations, however, the statistical significance within these percent differences was able to detect relatively small relationships. The widest differences were found between the high level violations and absconding. Proportions of general high-level violations ($d=0.03$, 95% CI [0.03, 0.05]) and sex related violations ($d=0.08$, 95% CI [0.07, 0.09]) were greater for men, while women held higher proportions of drug related violations ($d=0.08$, 95% CI [0.07, 0.09]). Similarly, most sanctions given were significantly different with the exception of the use of ERP. Jail was found to be given substantially more often to men than women ($d=0.06$, 95% CI [0.04, 0.07]), while women were more likely to receive verbal reprimands than men ($d=0.04$, 95% CI [0.03, 0.06]).

The remaining analyses were conducted with the data in wide format, with each individual represented only once ($n=18,886$). Cases were collapsed to this format by identifying the first sanction event and the highest violation associated with that event. For the sake of space, Table 2 summarizes the pre- and post-match sample descriptives on those measures that remained over a 15% bias (i.e., standardized percent difference) after the match, as well as measures of balance.[13] The only exceptions to reporting those covariates over the threshold are the demographic categories of age and race (dichotomized as nonwhite=1 and white=0), and the first violation-sanction event. These are presented as basic descriptives and were determined to be important in conceptualizing the nature of the sample.

As depicted in the model balance summary of Table 2, there were several notable differences between men and women before the match, including race, static risk, and criminogenic need measures. Approximately 81% of the covariates yielded a statistically significant difference. The majority of the measures possessed greater than 10% bias, and 23% of them were beyond 20% bias. Among the static risk and criminogenic needs with the greatest difference (i.e., over 20% bias), included those related to sex crimes, crimes related to weapon use/possession, employment and family related needs, as well as aggressive behaviors and attitudes. Most of these differences described a greater proportion of men possessed the criminogenic measures. Women made up the larger proportion among measures of misdemeanor sex crimes

16

Table 2. Bivariate descriptives pre- and post-match.

		Pre-match			Post-match		
		%Men	%Women	%Bias	%Men	%Women	%Bias
Model balance summary							
Percent significant differences			77.9			30.7	
Mean standardized percent difference or bias			12.3			4.4	
Maximum percent bias			54.1			27.2	
Percent with bias over 20			20.1			1.1	
Percent with bias over 10			47.2			7.8	
Area under the curve (AUC)			0.89			0.52	
Domain/measure (337 total measures compared)		%Men	%Women	%Bias	%Men	%Women	%Bias
Demographics	Sample size	18,886	3,220		3,197	3,197	
Age at supervision start	30–39	19.1	***22.1	7.4	28.7	30.9	4.8
	20–29	31.2	30.5	1.5	39.4	39.1	0.6
Non-white		29.3	***23.2	13.9	24.2	23.3	2.1
Adult criminal history							
Felony conviction history	Sex offense: One	5.4	***0.8	26.8	4.8	***1.1	22.0
	Two or more	4.4	***0.4	26.4	4.1	***0.5	24.2
Misdemeanor history	Sex offense: Two or more	0.4	***3.5	22.6	1.0	***3.5	16.9
Employment							
Employment barriers	Child care issues	0.6	***5.4	28.4	2.3	***5.4	16.2
Substance use/abuse							
Support for drug use/abuse	Prostitution	0.3	***5.6	31.7	0.8	***5.5	27.2

(continued)

Table 2. (continued)

Domain/measure (337 total measures compared)		%Men	%Women	%Bias	%Men	%Women	%Bias
Aggression							
Motivation for aggression	Power, dominance, or control	23.7	***5.3	54.1	9.7	***5.4	16.3
Attitudes/behaviors							
Motivation for criminal behavior during offender's lifetime							
Retaliation, vengeance		17.9	***3.0	50.2	6.5	***3.0	16.5
Money or material gain		9.5	***0.9	39.5	4.9	***0.9	24.0
First violation							
Any low-level violation		67.8	***71.4	7.8	72.8	71.4	3.1
Any high-Level violation		5.5	***2.3	16.6	3.4	**2.3	6.6
Abscond		14.8	***17.5	7.3	16.4	17.5	2.9
Number of violations at first sanction							
1 to 3 violations		60.8	*58.9	3.9	61.2	58.8	4.9
4 or more violations		39.2	*41.1	3.9	38.9	41.2	4.7
First sanction event							
Discretionary verbal reprimand		40.7	***49.7	18.2	46.9	*49.6	5.4
Enhance, reparations, or partial confine		5.2	*6.2	4.3	6.7	6.2	2.0
Targeted intervention/treatment		8.7	9.7	3.5	8.9	9.7	2.8
Jail		44.4	***33.3	22.9	36.4	*33.4	6.3
Prison		1.1	1.1	0.0	1.2	1.1	0.9

Note. $*p < .05$, $**p < .01$, $***p < .001$.

17

(e.g., prostitution to support drug use), minimal desire to continue education, economic marginalization. Economic marginalization examples included being impoverished (46% of women had a monthly income under $1,000 compared to 32% of men), lacking full-time employment (7% of women were fully employed compared to 19% of men), employment barriers involving child care and chemical dependency (28% of women with dependency issues versus 20% of men), relying public healthcare (33% of women to 14% of men), and relying on financial welfare services (30% of women rely on public assistance to 16% of men). Additionally, women had a greater percentage of cases with some type of formal mental health diagnosis (56%–34% of men), as well as currently relying on mental health medication (25%–11% of men). The widest pre-match difference involved the use of jail as a sanction (see Table 3), which was a 60% bias, with the overwhelming majority applied to men.[14]

Following the pre-match examination, the propensity scores were calculated to predict the likelihood a case was a woman based on the significantly different covariates noted above. It is important to note that we did not use all of the comparable covariates in the conditioning of the propensity score; the so called "kitchen sink" approach. Rather, only those that possessed a significant difference were included in the conditioning of the propensity score, which accounted for 264 covariates.[15] As shown in the model balance summary of Table 2, the propensity score yielded a strong AUC value (0.89) suggesting that the score successfully identified women using the measured covariates with 89% accuracy, prior to the match.

Post-Match Analyses

The bottom half of Table 2 shows the post-match balancing proportions between men and women. With strong common support, the greedy procedure yielded a nearest-neighbor match within the caliper (0.04, calculated via 0.25 multiplied by the standard deviation of the propensity score; see Rosenbaum & Rubin, 1985) for 3,197 women, leaving only 1% (23 cases) without a match.[16] Additionally, the AUC of the post-match dataset (0.52) indicated the match was successful in balancing the covariates between men and women. Aside from the AUC statistic, other measures suggested the match was successful. The average standardized percent bias dropped from 12.3 to 4.4, and the number of covariates with more than 10% and 20% bias was substantially reduced; of which the latter consisted of 1% of all covariates. Those covariates that remained above the 20% bias threshold, included using prostitution to support drug use, past sex offenses, and money or material gain being a primary motivating factor for criminality. Three other

Table 3. Breakdown of violation and outcome measures pre- and post-match.

Second violation-sanction experience	Pre-match			Post-match		
	%Men	%Women	d	%Men	%Women	d
Sample size	18,886	3,220		3,197	3,197	
Second violation						
Any low-level violation	52.8	**56.0	0.07	56.7	56.1	−0.01
Any high-level violation	4.0	***2.1	−0.37	3.1	*2.1	−0.22
Abscond	11.3	11.5	0.01	11.9	11.5	−0.02
Rearrest (misdemeanor or felony)	6.8	***4.7	−0.22	5.9	*4.7	−0.13
Number of violations at second sanction						
1–3 violations	46.1	***42.6	−0.08	46.7	**42.5	−0.09
4 or more violations	34.8	35.1	0.01	33.9	35.2	0.03
Second sanction event						
Discretionary verbal reprimand	13.9	**15.9	0.09	14.3	16.0	0.07
Enhance, reparations, or partial confine	4.0	**5.3	0.16	4.6	5.3	0.08
Targeted intervention / treatment	6.4	7.3	0.08	7.0	7.2	0.02
Jail	40.9	***33.5	−0.18	39.4	***33.6	−0.14
Prison	1.0	0.8	0.12	1.2	0.8	−0.23

(continued)

19

Table 3. (continued)

Subsequent recidivism	%Men	%Women	d	%Men	%Women	d
Technical violations						
Any low-level violation	34.6	35.6	0.02	35.9	35.6	-0.01
Any high-level violation	2.5	***1.0	-0.51	2.0	**1.0	-0.39
Abscond	9.0	8.2	-0.06	8.9	8.3	-0.04
Rearrest (misdemeanor or felony)	4.4	***2.8	-0.26	4.2	**2.8	-0.23
Reincarceration						
Jail	31.1	***27.4	-0.10	30.2	*27.5	-0.07
Prison	24.7	***15.5	-32	21.3	***15.6	-0.21
Reconviction (misdemeanor or felony)	25.6	**28.1	0.07	28.0	28.2	0.01
Any return to custody	42.5	***35.1	-0.17	40.9	***35.2	-0.13

Note. $*p < .05$, $**p < .01$, $***p < .001$.

covariates remained somewhat close to the threshold including employment barriers involving childcare issues (16.2), motivation for aggression being power, dominance or control (16.3), and criminal motivation being retaliation or vengeance (16.5). Table 2 also reports the breakdown of pre- and post-match proportions of the first violation and sanction experienced by both men and women. It is important to note that the factors that largely determine the type of sanction (i.e., number of violations and violation type) were matched tightly; none had a standardized percent bias over 6.6. Similarly, the type of sanction was also closely matched between men and women, suggesting that the post-match sample is prepared to answer the primary research question.

Deeming the match to be an overall success, we conducted a bivariate examination of the outcome measures using chi-square tests. Table 3 reports the proportions across the second violation-sanction event and subsequent recidivism after the second event, along with the corresponding Cohen's d value. Cohen's d is coded such that negative values indicate a greater proportion of the outcome of interest among men compared to women offenders, and positive values indicate a greater proportion among women. Four measures yielded a significant difference in the matched sample for the second violation-sanction event as well as in the subsequent recidivistic event. Any high-level technical violation, absconding, and reincarceration were significantly different for jail in the second event, and for prison in the subsequent event. In terms of effect size, the more sizable differences were among subsequent prison reincarceration ($d=-.21$, 95% CI [−0.27, −0.15]) and subsequent high-level violations ($d=-.39$, 95% CI [−0.59, −0.19]). This suggests there is a difference in how men and women behave following their first and section sanctioning events.

Logistic regression. To isolate the effects of gender while accounting for remaining differences across the unbalanced covariates and sanctions experienced, a series of logistic regressions were conducted. The post-match, significantly different outcome measures were regressed onto gender without (baseline) and with (main effect) the unbalanced covariates and sanction information. We focused our multivariate analyses on the measures of high-level violation, rearrest, return to prison, and any return to custody. Considering returning to jail yielded a small effect size, its relationship to gender can be accounted for in the any return to custody measure. Returns to prison were left separate as it had a larger effect size immediately after the match. High-level violations did not have enough observations to fit the models properly. Taking absconding as it is generally classified, a high-level violation, the two measures were combined for the purposes of the multivariate analysis.

Table 4 presents the findings of the logistic regressions with high-level violations/absconding, return to prison, and any return to custody following the first violation-sanction event. While the unbalanced covariates were controlled for in these models and are represented in the model balance statistics, this information is not shown in the table for the sake of parsimony and space.[17] The baseline model incorporates gender as the sole predictor of each outcome. The first set of main effects models includes sanction type experienced for both the first and second violation as well as unbalanced covariates. The second sanction experienced is not reported in this set because those offenders who experienced only one sanction (the first sanction) were set as the reference group along with reprimands. This allowed for those who received a second sanction to be included in the models and their type to still be held constant. The second set of main effects models (shown in Table 5) examines the same measures but only using those cases who experienced a second sanction ($n=4,135$) with reprimand set as the reference group. Separating these cases provided a more accurate effect of each type of sanction on the recidivism measures.

The fit indices shown in Tables 4 and 5 indicated that all final models were an improvement over the intercept models. Post-match, baseline models revealed gender was a significant predictor of all outcome measures examined. However, the predictive strength of gender was weakened when the other covariates were included in the model. Main effects models examining the full sample indicated that, all else held constant, gender was a significant predictor ($\alpha=0.05$) of returning to prison. Although no longer significant at the 0.05 level within the main effects models of rearrest and high-level and absconding violations, p values of the odds ratios (OR) suggest there was between a 6% (rearrest) and 25% (high violation and abscond) likelihood of achieving the respective coefficients (greater odds for men to have a high-level violation [10%] and to be rearrested [24%]), assuming the null hypothesis is true (i.e., no relationship between gender and recidivism events).[18] That said, gender remained a significant predictor for each of the other outcomes indicating that men possessed greater odds than women of being returned to prison (33%), and returned to any custody (17%).

In examining the other covariates, it appears there are several reoccurring and remarkable aspects. Holding no sanction/reprimand as the reference group, receiving jail as a first sanction appeared to consistently increase the odds of all outcomes. The effects subside when jail is the second sanction received, although it still increased the likelihood of a return to prison and any custody. Prison sanctions appear to have little to no effect on supervision failures with the exception of instances where it is the first sanction. In such cases, prison decreased the likelihood of returning to custody and increased

Table 4. Logistic regression predicting subsequent recidivism measures accounting for the first violation-sanction event.

	High viol. or abscond			Rearrest			Return to prison			Any return to custody		
Baseline models (n=6,394)	OR	[95%CI]	p	OR	[95%CI]	p	OR	[95%CI]	p	OR	[95%CI]	p
Men=0; Women=1	0.84	[0.71–0.99]	.034	0.65	[0.49–0.85]	.002	0.68	[0.60–0.78]	<.001	0.79	[0.71–0.87]	<.001
Constant	0.12	[0.11–0.14]	<.001	0.04	[0.04–0.05]	<.001	0.27	[0.25–0.30]	<.001	0.69	[0.64–0.74]	<.001
Main effects models, full sample (n=6,394)												
Men=0; Women=1	0.90	[0.76–1.07]	.247	0.76	[0.58–1.02]	.064	0.67	[0.58–0.78]	.001	0.83	[0.73–0.93]	.002
First sanction experienced												
Enhanced/payment/partial confinement	0.45	[0.27–0.75]	.002	.99	[.55–1.80]	.978	0.81	[.57–1.17]	.262	0.81	[0.63–1.05]	.115
Treatment programming	0.92	[0.65–1.29]	.617	1.44	[.89–2.33]	.134	0.72	[.52–1.00]	.052	0.79	[0.63–1.00]	.048
Jail sanction	1.40	[1.16–1.68]	<.001	1.38	[1.01–1.88]	.040	1.59	[1.37–1.85]	<.001	1.52	[1.33–1.73]	<.001
Prison sanction	0.20	[0.03–1.44]	.109	.63	[.09–4.65]	.653	0.71	[.30–1.69]	.436	0.42	[0.20–0.87]	.020
Constant	0.06	[0.03–0.11]	<.001	.04	[.02–0.13]	<.001	0.81	[.57–1.17]	.262	0.08	[0.05–0.14]	<.001
Likelihood-ratio χ^2	***331.55 (34 df)			***137.52 (34 df)			****1006.60 (34 df)			***1792.19 (34 df)		
McFadden's R^2	0.08			0.07			0.17			0.21		
BIC'	−33.60			160.43			−708.65			−1494.25		

Note. $*p<.05$, $**p<.01$, $***p<.001$.

23

Table 5. Logistic regression predicting subsequent recidivism measures accounting for the second violation-sanction event.

Main effects models, sample with second sanction (n = 4,135)	High vio. or abscond			Rearrest			Return to prison			Any return to custody		
	OR	[95%CI]	p	OR	[95%CI]	p	OR	[95%CI]	p	OR	[95%CI]	p
Men = 0; Women = 1	0.90	[0.75–1.07]	.242	0.75	[0.56–0.99]	.046	0.65	[0.56–0.76]	<.001	0.77	[0.67–0.89]	<.001
First sanction experienced												
Enhanced/payment/partial confinement	0.38	[0.23–0.63]	<.001	0.87	[0.48–1.57]	.634	0.69	[0.48–0.98]	.041	0.61	[0.46–0.79]	<.001
Treatment programming	0.81	[0.57–1.15]	.237	1.28	[0.79–2.07]	.319	0.63	[0.45–0.89]	.008	0.63	[0.49–0.80]	<.001
Jail sanction	1.23	[1.02–1.48]	.033	1.19	[0.87–1.63]	.264	1.33	[1.14–1.57]	<.001	1.20	[1.03–1.39]	.020
Prison sanction	0.25	[0.03–1.87]	.176	0.91	[0.12–7.08]	.929	0.92	[0.35–2.40]	.863	0.46	[0.19–1.07]	.070
Second sanction experienced												
Enhanced/payment/partial confinement	0.66	[0.46–0.94]	.023	0.82	[0.49–1.36]	.437	0.57	[0.39–0.83]	.004	0.61	[0.48–0.77]	<.001
Treatment programming	0.54	[0.35–0.85]	.007	0.75	[0.41–1.37]	.356	0.57	[0.41–0.79]	.001	0.46	[0.35–0.61]	<.001
Jail sanction	1.06	[0.86–1.31]	.581	0.82	[0.59–1.14]	.227	1.25	[1.05–1.50]	.013	1.43	[1.21–1.68]	<.001
Prison sanction	0.63	[0.28–1.44]	.275	0.39	[0.09–1.67]	.206	1.71	[0.99–2.97]	.056	0.60	[0.35–1.02]	.056
Constant	0.38	[0.23–0.63]	<.001	0.87	[0.48–1.57]	.634	0.69	[0.48–0.98]	.041	0.61	[0.46–0.79]	<.001
Likelihood-ratio χ^2	***103.73 (34 df)			***75.5 (34 df)			***504.68 (34 df)			***541.55 (34 df)		
McFadden's R^2	0.03			0.04			0.10			0.09		
BIC'	179.40			207.63			−221.55			−231.42		

*p < .05, **p < .01, ***p < .001.

the likelihood to return to prison if it is the second sanction. Treatment and the ERP option both decreased the likelihood of returning to custody, particularly for those who experienced a second sanction (Table 5). For those who received a second sanction, both treatment and the ERP option are attributed to possessing 39% lower odds to have a subsequent high/absconding violation and 54% lower odds of a return to custody. Additionally, the ERP option reduced the odds of high/absconding violations by between 34% and 62% for both the full sample and those experiencing a second sanction event.

With the focus on how gender influences the effects of supervision sanctions, we tested the interaction of the two measures.[19] Although most of the interactions were neither statistically significant nor substantively different between the genders, there are a few noteworthy findings. First, in the model predicting the likelihood of a subsequent high-level/abscond violation, people on supervision of either gender receiving an ERP as their second sanction possessed a baseline 9.1% chance of committing the predicted outcome (logistic model interaction $p = .049$). This statistically significant interaction effect suggests that men who received a second sanction had a 6% chance of committing another high-level/abscond violation, whereas women receiving a second sanction had a 12% chance. Second, when prison was used as a first sanction, the chances of the average released individual (man or woman) to ultimately return to prison was 11.3% and 19.6% for any return to custody. The interaction effects indicated that for an otherwise average man on supervision, a first sanction of prison for a technical violation resulted in approximately a 5% chance of a subsequent return to prison and return to any custody. In contrast, it resulted in a 17.7% and 34.9% chance for the average woman to return to prison (interaction $p = .086$) and any custody (interaction $p = .009$), respectively. For those who received second sanctions, no interactions of sanction type and gender were found to be significant. However, second sanctions of jail and prison continued to yield the highest likelihood to return for both men and women. The average chance to return to prison ranged from a 31.4% (after jail as a second sanction) to 37.8% chance (after receiving prison as a second sanction). Notably among the interactions, the lowest chances to return to prison were after receiving ERP or treatment for a second sanction (18.2% chance each, respectively). This was even lower for women who received ERP or treatment for their second sanction (15% each), compared to men who received ERP (21.6%) or treatment (21.1%).

Survival analysis. To further account for the variation in release dates and the rates at which men and women may differ in their failure of supervision, we conducted Cox proportional hazard regressions for each supervision outcome. These survival analyses aimed to account for multiple covariates in the

prediction of time-to-event, rate, and risk of outcomes (Hosmer et al., 2008). For this study, special attention was given to the rate at which women fail supervision at a given time, compared to the matched comparison group of men. Moreover, we examined the ability of gender to predict the relative risk of failure given a particular sanction experience. Procedures used in conducting the survival analysis were similar to that of the logistic regression (i.e., examining the baseline model with just gender as a covariate [women = 1], then adding unbalanced covariates to assess any change, and finally testing interactions).

In examining failure rates between men and women, Kaplan-Meier functions and log-rank tests of equality revealed only those outcomes with significant differences from the post-match analysis yielded a significant difference in the rate of failure: High-level/abscond violations, rearrest, returns to prison, and any return to custody. High-level/abscond violations were generally committed at a lower rate than most other failures. Only about 2.3% of men and 1.3% of women committed a high-level violation within the first year following their second sanction. Beyond a year, the difference widens to statistical significance ($\chi^2 = 11.0$, $p < .001$). Similar differences ($\chi^2 = 9.9$, $p = .002$) were detected in rearrests at one year, as men were rearrested at a slightly higher rate (4.0%) than women (2.7%). Larger differences were observed in returns to prison, as men were returned at a significantly higher rate (12.4%) than women (9.7%) in the first year ($\chi^2 = 37.3$, $p < .001$). The difference in rates grew to 23.1% for men and 16.6% for women by the end of the study period. Lastly, within the first year, men were returned to some form of custody (i.e., arrest, jail, or prison) at a significantly higher rate (38.5%) than women (33.9%; $\chi^2 = 23.2$, $p < .001$). This rate leveled out after 27 months, but the difference between the genders remained, ultimately ending with 41.5% of men and 35.5% of women being returned to custody.

Upon conducting the Cox proportional hazards regression for each of the outcomes with differences in rates of failure, the main effects models confirmed the log-rank tests when introducing the unbalanced covariates. Although gender coefficients and its predictive strength fluctuated slightly as covariates were introduced, the main effects of gender generally remained a statistically significant predictor. After controlling for all unbalanced covariates and sanction types, gender was a significant predictor of failure rates for all outcomes. Women (coded as 1) had a significantly lower propensity to commit high-level/abscond violations (hazard ratio [HR] = 0.82, $p = .017$, 95% CI [0.70, 0.96]), be rearrested (HR = 0.76, $p < .054$, 95% CI [0.58, 1.00]), returned to prison (HR = 0.74, $p < .001$, 95% CI [0.66, 0.84]), and have any return to custody (HR = 0.89, $p = .006$, 95% CI [0.82, 0.97]).

Most importantly are the effect of gender in relation to specific sanctions. Overall, each sanction type significantly predicted outcomes in roughly the same manner as found in the binary logistic regression, although the strength of these relationships varied across outcomes. In order to identify the effects of gender given a specific type of sanction experienced at both the first and second violation-sanction event, we calculated the marginal effects of each interaction model. Table 6 provides the marginal interaction effects on outcomes for women at each sanction type compared to the men's baseline hazard ratio. Negative coefficients in Table 6 indicated that women have a lower likelihood than men to fail by the column outcome. Gender presents little to no effect as it relates to the types of sanctions experienced, with some exceptions. Only a few effects reached the Neyman-Pearson levels of statistical significance and only one sanction type (i.e., prison as a first sanction) indicated women had a significantly higher likelihood to fail.

Sensitivity assessment. To test the degree to which hidden bias (i.e., influence of unobserved variables) may have impacted our models and overall findings we conducted a sensitivity analysis. Particularly in 1 to 1 matches, this process estimates how important unobserved measures can be for the findings by determining "what the unmeasured covariate would have to be like to alter the conclusions of the study" (Rosenbaum, 2005, p. 1809). Specifically, we used the Stata command *mhbounds* (Becker & Caliendo, 2007) that uses a Mantel-Haenszel test to determine the magnitude that hidden bias must account for in order to change the outcome, designated by the measure Γ (Gamma; Rosenbaum, 2002, p. 107). The higher the Gamma, the less sensitive results are to hidden bias. Sensitivity tests of each outcome variable suggested that the findings of this study are somewhat sensitive to hidden bias. Assuming an overestimation of the treatment effects, the highest significant Γ was found for high-level violations ($\Gamma = 2.3$, not very sensitive), while the other three outcome measures that yielded a difference between the genders ranged in Γ from 1.2 to 1.3 (somewhat sensitive). This suggests that if unobserved measures were introduced into the model, they would only need to increase the odds of a successful prediction of gender by a factor of at least 1.2 and at most 2.3 before the results might change.

Discussion

This study provides important insight about the role that gender plays in offender responses to community supervision sanctions. Our findings suggest there is a difference in the violation behavior of men and women offenders. Women were shown here to commit more low-level, drug-related

Table 6. Survival model marginal effects of failure and gender at certain sanction types.

Woman (1) at	High vio. or abscond			Rearrest			Return to prison			Any return to custody		
	ME	[95%CI]	p	ME	[95%CI]	p	ME	[95%CI]	p	ME	[95%CI]	p
Men HR[a]	1.15	[1.48–0.9]	.270	1.28	[1.98–0.83]	.270	1.41	[1.71–1.16]	.270	1.16	[1.31–1.02]	.026
First sanction												
Reprimand	−0.41	[−1.10–0.29]	.255	−0.34	[−0.91–0.24]	.250	−0.66	[−1.04–0.28]	.001	−0.35	[−0.65–0.05]	.021
Treatment	−1.84	[−4.02–0.35]	.100	−0.79	[−2.41–0.83]	.337	−0.53	[−1.49–0.43]	.281	−0.28	[−0.99–0.43]	.441
ERP	−1.37	[−4.04–1.30]	.313	−1.20	[−3.05–0.64]	.200	−0.05	[−.78–0.69]	.899	−0.26	[−0.86–0.34]	.398
Jail	−0.53	[−1.18–0.13]	.115	−0.34	[−1.1–0.43]	.387	−0.94	[−1.54–0.34]	.002	−0.31	[−0.69–0.06]	.099
Prison	1.76	[−1.81–5.33]	.333	1.75	[−1.98–5.48]	.357	2.31	[−0.47–5.08]	.104	2.37	[.53–4.21]	.011
Men HR[a]	1.41	[1.96–1.01]	.042	1.50	[2.53–0.88]	.134	1.43	[1.85–1.11]	.005	1.13	[1.35–0.96]	.150
Second sanction												
Reprimand	−0.24	[−0.46–0.02]	.032	−0.25	[−0.57–0.07]	.120	−0.28	[−0.46–0.09]	.003	−0.09	[−0.22–0.03]	.131
Treatment	0.56	[−0.14–1.26]	.119	−0.40	[−1.02–0.22]	.209	−0.13	[−0.39–0.14]	.361	−0.18	[−0.35–0.01]	.043
ERP	−0.48	[−1.07–0.12]	.116	−0.08	[−0.53–0.38]	.742	−0.17	[−0.39–0.06]	.152	−0.14	[−0.30–0.02]	.081
Jail	−0.09	[−0.22–0.03]	.148	−0.14	[−0.35–0.08]	.214	−0.28	[−0.45–0.10]	.002	−0.09	[−0.18–0.01]	.066
Prison	−0.63	[−1.42–0.17]	.121	0.71	[−0.42–1.84]	.219	−0.60	[−1.96–0.77]	.392	−0.12	[−0.61–0.37]	.642

aBaseline hazard ratio for men in the given model, from which the marginal effect for women is based.

violations and abscond from supervision slightly more frequently, while men committed more high-level violations. There was also a difference found in the types of sanctions given to men and women by their supervising parole officers. Women received more discretionary no sanctions/verbal reprimands, while men received more jail sanctions. Finally, there were differences found in the supervision outcomes between men and women who technically violated and received a sanction. Multiple analyses suggest that gender is a meaningful factor in how violators respond to the sanctions received, although the strength and direction of this association were dependent on the outcome of interest (e.g., absconding versus returns to prison).

These findings must be interpreted in the context of the study's limitations. First, there are tradeoffs in using a sample of adults on supervision who have been sanctioned at least once. It is possible that the mere threat of sanctioning someone on supervision might influence men and women differently. With this in mind, it was our aim to attempt to control for this issue by including the instances in which the parole officer captured a verbal reprimand, and thereby accounting for such a threat and response. Also, while the choice of using the first violation-sanction event allowed for a methodological standardization of sanction experiences upon release, it also limited our ability to speak to offenders' entire supervision experience. It is possible that a person's first violation-sanction event was less meaningful than the second or third experience. In which case, this analysis does not capture such an instance very well. However, our time-to-event analyses and other studies have suggested (e.g., Grattet & Lin, 2016) that the longer the individual remains in the community, the greater the likelihood they will remain crime-free. One particular strength of using the first sanctioned event is we are able to examine those who received a reprimand or discretionary no-sanction instead of a formal jail stay or ERP. This group is commonly overlooked in the literature and this study provides a brief depiction of how these individuals respond in comparison to others who receive different sanctions. Examining offender responses in greater detail in future research may provide opportunities to further explore the importance of less stringent graduated sanctions.

Despite these limitations, the findings of this study contribute to the understudied knowledge base about violations, gender, and sanctions. First, the observed differences in sanctions applied to men and women raise concern about patterned responses that are not necessarily guided by the graduated sanctions. While some of these patterns appear to coincide with the differences in violation types (e.g., women had more drug violations and received more inpatient treatment responses), others appear to hold the potential for gendered patterns in how parole officers may use or view sanctions. For

instance, women received a significantly larger proportion of less punitive sanctions (i.e., verbal reprimands), and men received larger proportions of more punitive sanctions (i.e., reparations/fees/fines, or jail time). While the strength of these relationships must be further explicated by future research, the potential policy implications for jurisdictions using similar graduated sanctions is to further structure discretion and guide officers in using fewer punitive sanctions and a greater effort in addressing criminogenic needs highlighted by the violating behavior.

Second, recall that the gender-responsive perspective argues women offenders possess unique attributes and criminogenic needs that should be addressed differently than men, whereas the principles of effective intervention maintains there are no substantive differences in the importance of the criminogenic needs of men and women offenders; instead, any differences are indicative of responsivity factors rather than needs. In this study, we found evidence of differences in the attributes between men and women, which we interpret as partial support for both perspectives. Although a third of the covariates remained significantly different after the match, we argue that these differences provide a partial answer to our research question in noting an unmatchable difference between men and women on criminogenic needs. Such covariates (shown in Table 2) that remained different suggest distinct, key items that set men and women offenders apart. A brief examination of the covariates remaining over 10% bias following the match reveals a number of measures depicting gendered differences in relation to either socialization or systemic approaches.

Socialization measures included those related to the traditionally gendered positions in criminal activity (e.g., using prostitution to support drug use), assuming a traditionally gendered social role (e.g., primary care-provider for minor children), or economic marginality (e.g., being unemployed but able to work). Systemic measures involved those that depict a common application of the criminal justice or health and human services systems that play into gendered roles (i.e., the system and its actors "doing gender"). For instance, such measures include requirements of child support expected of the father and less custody given to the father over minor children. This is indicative of how courts may award custody more often to mothers than fathers as a general rule, but has little basis in empirical evidence to justify it. Such differences were reiterated in the logistic regression predictors of various outcomes, especially those predicting returns to prison. Furthermore, the ability of the unbalanced covariates to predict technical violations and returns to prison similarly is consistent with the literature in suggesting there may exist a qualitative difference between technical violations and new crimes, especially as they relate to men and women (see Campbell, 2014; Hamilton & Campbell, 2013).

On the other hand, the majority of the 337 factors did not yield a percent bias over 20%. As a result, from a gender-neutral perspective one may argue that unbalanced factors are not worth highlighting. In this sense, it could be that the overall covariate differences are not large enough to be meaningful, and hence, are not an indication of gendered differences within criminogenic needs domains. Similarly, those covariates that are worth noting (e.g., prostitution for drugs) can be addressed using the same service approaches as men; using work release employment services coupled with cognitive-behavioral treatment that helps to change anti-social attitudes/beliefs, and simultaneously addressing substance abuse or chemical dependency issues in relapse prevention strategies. Furthermore, gender-neutral advocates may point to the diminished significance of unbalanced covariates in the main effects logistic regression models. This would suggest that the ability of such factors to predict the outcome is often washed out by more static measures, which are widely viewed as gender-neutral. Whether the differences found here operate as needs or responsivity factors remains an area for further academic exploration.

Regarding the outcomes and considering that several criminogenic needs are accounted for in both the propensity scores and in the final multivariate models, it appears that there may be a gender difference worth exploring further. Even after accounting for what some would call universal criminogenic needs, key differences still emerge in how men and women behave following a sanction for a technical violation. However, it is again important to note that the Cohen's *d* statistics, regression odds and hazard ratios generally indicated a relatively small effect size difference between men and women on most outcomes. That being said, some effects are remarkable, such as using the ERP options. Women receiving ERP as their second sanction possessed a likelihood of having a high/abscond violation that was twice that of the men. This finding lends support to a potential for women to fail supervision when more formal controls are put in place (i.e., condition enhancements, fines, and partial confinement; Blackwell, 2000; Hamilton & Campbell, 2013). This is consistent with studies by Morash et al. (2019) who noted that some women may respond differently to punitive sanctions depending on when and how they are applied. Similarly, while treatment holds the potential to reduce supervision failure, the lack of definitive and consistent effects for treatment suggests that the application of treatment and its effectiveness is far more nuanced than what could be examined here. Moreover, as Morash et al. (2019) highlighted, there may be a qualitative difference in the types of women, and potentially men, on supervision who respond to treatment.

Future research should seek to examine how men and women differ in their response to combinations of specific sanction types. For example, future studies could assess whether men and women respond in a similar way to the use of jail as a sanction when it is coupled with a treatment program. This is particularly pertinent given the latest push for the use of jail as a swift-and-certain sanction in many jurisdictions (see Hawken, 2010; Kleiman et al., 2014). A final and similar point to note regarding the findings is the inability of confinement (particularly jail) to reduce the likelihood of both genders to recidivate by any measure. Considering that jail is by far the most frequently used sanction for both men and women, this promotes considerable concern about the expected effectiveness of confinement to reduce recidivism among technical violators. In a similar vein, although far less common, being sanctioned to prison (especially as a first sanction event) appears to be more damaging for women than men as women were more than twice as likely to return to any custody following such a sanction. Considering the context of debating perspectives in corrections, our findings supply evidence of differences between men and women, albeit small to modest. It is possible that employing a more gender-responsive approach (e.g., trauma informed care) to how parole officers address technical violations and sanctions may help improve offender outcomes beyond what was observed here – although, this remains an area in need of more research. Overall, the findings presented here provide a foundation from which future research can dive deeper into potential gender differences in determining when they occur, why they occur, their more specified magnitude, and what such differences might mean (if anything) for correctional practice and policy.

Declaration of Conflicting Interests

Funding

ORCID iD

Christopher M. Campbell https://orcid.org/0000-0002-5443-5523

Supplemental Material

Supplemental material for this article is available online.

Notes

1. The term technical violation is used here to refer to any noncriminal breach of the supervision conditions imposed on an offender under the auspices of community corrections. Due to the non-criminal nature of technical violations, there are many ways to examine and interpret such behavior as violations can relate to probation and parole officer discretion (Campbell, 2014). This study assumes that the correctional agencies and practitioners view technical violations as problematic behavior.

2. Although incarceration is discussed as a graduated sanction, it is distinct from other lower-level sanction types. The use of community-based sanctions, for example, suggests the offender can safely receive treatment services in society, whereas the use of jail or prison indicate the offender is not yet ready to be free of control (Campbell, 2017).

3. It is beyond the scope of the current study to make conclusions about the effectiveness or importance of gender-responsive programming, trainings, or tools. Rather, the study aim is to emphasize differences in how men and women respond to common/traditional sanctions seen in graduated sanctioning practices. This may provide a foundation for future research to estimate the effectiveness of gender responsive treatment.

4. The terms "parole" and "parolee" are no longer used in Washington. However, because the terms used instead, "community supervision" and "community corrections participant" (respectively), are conceptually similar in that they are conditional releases, the former are used for the sake of simplicity and general connection to the same process in most states. Parolees were used in this study rather than probationers because the former generally involves individuals who are at a higher risk for recidivism than the latter. If responses to sanctions do vary by one's gender, therefore, this difference should be most apparent among parolees.

5. As of late 2012, WADOC implemented swift-and-certain sanctioning which impacted the frequency of jail use as a sanction for technical violations (Hamilton, Campbell, et al., 2016). During this time, the WADOC had not adopted any gender-responsive trainings, tools, or programs. Since this time period, WADOC has engaged in gender-responsive approaches (e.g., implementation of the Women's Risk Needs Assessment, WADOC, 2014).

6. While the ONA was not validated to predict recidivism until 2015 in combination with the SRA measures (Hamilton et al., 2016), the SRA was validated in 2006 (Barnoski & Drake, 2007) using data on offenders supervised in Washington from 1986 to 2000 (construction sample of 308,423 observations) and from 2001 to 2002 (validation sample of 51,648 observations). The SRA was found to predict three types of felony recidivism with a moderate degree of accuracy and a validation-specific area under the curve values at 0.733 (property/violent), 0.732 (violent), and 0.742 (any felony). WADOC policy during this period dictated that ONA information was to be gathered via an interview that lasted about an hour. While the SRA was only collected at intake, the ONA was collected at multiple times during the WADOC cycle. The ONA was

collected at intake, up to four weeks prior to release to community supervision, and then reassessed periodically while on supervision. All case managers and community corrections officers conducting ONA interviews were trained in a formal academy, and while there are some quality assurance measures, no internal process was established for assessing ongoing reliability (for further information, see Hamilton et al., 2016).

7. Some missing data was still present in the individual SRA measures (less than 15%). Considering these items only made up a small percentage of the covariates used for the match, an iterative matching process was able to distribute the measures across the gender groups. This was not deemed to be an issue of concern for the findings considering the balance between the groups on both SRA and ONA items.

8. The full matching table (pre- and post-match bivariate) breaking down all 337 covariates is included as an appendix.

9. WADOC policy specifies a definition for high- and low-level violations that include the specified categories. Some special supervision cases (e.g., such as those supervised under the Drug Offender Sentencing Act) include exceptions to the operationalization of high-level violations. These were excluded as their supervision terms are specific to first-time drug offenders. Among the final eligible sample, there were no cases whose first violation was a rearrest. This was a rather rare occasion (shown in Tables 1 and 3) as most new crimes were not recorded as a technical violation and are instead reported to the district attorney as a new charge and thereby counted as recidivism. It is likely that among those people who were removed in the eligibility criteria there were several people who recidivated without ever recording a violation.

10. Unfortunately, treatment completion was not tracked by the WADOC.

11. According to WADOC officials (Harrison, 2017, personal communication), most confinement was not an option for community corrections officers to use directly. Rather, confinement was reserved as a sanction from hearing officers. If someone was returned to prison and then released again with enough time still on their sentence to be placed back out on community supervision, their subsequent behavior after the prison sanction was tracked.

12. It should be noted that this is not the official definition of recidivism according to the WADOC.

13. For more information on all covariates we refer readers to the appendix.

14. For more information on these differences we refer readers to the appendix.

15. Although it has been cautioned that covariates should not be used that could be effected by the treatment condition (Wooldridge, 2005), this concern does not seem applicable in this situation because gender is embedded to an equal degree of both the covariates and outcome measures. Additionally, considering PSM is a "data hungry" method (i.e., the more information, the more accurate the match), this number of covariates could be expected to supply a match to most, if not all of the women. Other research has employed PSM on such a large number of covariates, with similar expectations (e.g., Hamilton, Campbell, et al., 2016).

16. Upon examining the 1% unmatched, it became apparent that unique features of these cases were spread across a number of covariates that would minimize the likelihood of finding a close enough match within the specified caliper for the propensity score. In other words, these individuals were too diverse across the covariates to find an adequate match. That said, there are four covariates that 90-100% of the unmatched cases of women had in common with each other: All of them had a drug use problem, 90% had a specific history of methamphetamine use, all had a history of aggressive behavior, and all had a documented diagnosis of a mental health disorder.

17. Of all the unbalanced covariates, only those related to having children who were still minors and if the offender lived with his/her children were statistically significant at the 0.05 level for returns to prison and any return to custody models. Both measures suggested that offenders who have or are living with their children were 30% less likely to return to prison or any custody.

18. It is recognized that this interpretation is not in concert with conventional sociological and criminological interpretations of Neyman-Pearson's null hypothesis testing in setting an alpha level and interpreting coefficients appropriately. Instead, Fisher's approach in viewing the *p* values as a measure of strength in the regression coefficients against an assumed null hypothesis was used. Apart from providing a more meaningful analysis, interpreting the strength of marginal differences between men and women is critical given the nature of the aforementioned gender debate. That said, we provide the asterisks in our tables for quick reference. Additionally, we applied the Neyman-Pearson's method in the bivariate analyses for ease in interpretation of preliminary analyses to the main effects models. For a recent discussion and summary of this debate, we refer readers to Cumming and Calin-Jageman (2016).

19. The full models are not shown, but available upon request. All findings reported here are the result of the marginal effects from the tested full models.

References

Andrews, D. A., & Dowden, C. (1999). A meta-analytic investigation into effective correctional intervention for female offenders. In *Forum on corrections research* (Vol. 11, No. 3, pp. 18–21), Correctional Service of Canada.

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both 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.

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

Barker, V. (2011). Decarceration. *Criminology & Public Policy*, *10*(2), 283–286.

Barnes, G. C., Ahlman, L., Gill, C., Sherman, L. W., Kurtz, E., & Malvestuto, R. (2010). Low-intensity community supervision for low-risk offenders: A randomized, controlled trial. *Journal of Experimental Criminology*, *6*(2), 159–189.

Barnoski, R., & Drake, E. K. (2007). *Washington's offender accountability act: Department of corrections' static risk assessment*. Washington State Institute for Public Policy.

Becker, S. O., & Caliendo, M. (2007). Sensitivity analysis for average treatment effects. *Stata Journal*, *7*(1), 71–83.

Blackwell, B. (2000). Perceived sanction threats, gender, and crime. *Criminology*, *38*(2), 439–488.

Bloom, B., Owen, B., & Covington, S. (2004). Women offenders and the gendered effects of public policy. *Review of Policy Research*, *21*(1), 31–48.

Bogue, B., Campbell, N., Clawson, E., Faust, D., Florio, K., Joplin, L., Keiser, G., Wasson, B., & Woodward, W. (2004). *Implementing evidence-based practice in community corrections: The principles of effective intervention* (No. 019342). National Institute of Corrections, Community Corrections Division; Crime and Justice Institute.

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). Routledge.

Bonta, J., Rugge, T., Scott, T., Bourgon, G., & Yessine, A. K. (2008). Exploring the black box of community supervision. *Journal of Offender Rehabilitation*, *47*(3), 248–270.

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.

Bucklen, K. B. (2014). *The impact of returning technical parole violators to prison: A deterrent, null, or criminogenic effect* (Dissertation). University of Maryland, College Park.

Campbell, C. M. (2014). It's not technically a crime investigating the relationship between technical violations and new crime. *Criminal Justice Policy Review*, *27*(7), 643–667.

Campbell, C. M. (2015). *Dooming failure: The impact, utility, and practice of revoking conditional release* (Dissertation). Washington State University.

Campbell, C. M., & Labrecque, R. M. (2018). Panacea or poison: Can propensity score modeling (PSM) methods replicate the results from randomized control trials (RCTs)?" (Award Number: 2016-R2-CX-0030). Final report to the National Institute of Justice.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Cook, N. R. (2008). Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry*, *54*(1), 17–23.

Cumming, G., & Calin-Jageman, R. (2016). *Introduction to the new statistics: Estimation, open science, and beyond* (Reprint ed.). Routledge.

Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, *42*(4), 449–467.

Drake, E. K., & Aos, S. (2012). *Confinement for technical violations of community supervision: Is there an effect on felony recidivism?* (No. 12-07-1201). Washington State Institute for Public Policy.

Durose, M. R., Cooper, A. D., & Snyder, H. N. (2014). *Recidivism of prisoners released in 30 states in 2005: Patterns from 2005–2010* (Special Report No. NCJ 244205). Bureau of Justice Statistics.

Erwin, B. S. (1986). Final report of the georgia intensive probation supervision project. Department of Corrections.

Finn, M. A. & Muirhead-Steves, S. (2002). The effectiveness of electronic monitoring with violent male parolees. *Justice Quarterly*, *19*(2), 293–312.

Franklin, T. W. (2015). Race and ethnicity effects in federal sentencing: A propensity score analysis. *Justice Quarterly*, *32*(4), 653–679.

Gendreau, P., French, S., & Gionet, A. (2004). What works (what doesn't): The principles of effective correctional treatment. *Journal of Community Corrections*, *13*(3), 4–6.

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.

Grattet, R., & Lin, J. (2016). Supervision intensity and parole outcomes: A competing risks approach to criminal and technical parole violations. *Justice Quarterly*, *33*(4), 565–583.

Grattet, R., Petersilia, J., Lin, J., & Beckman, M. (2009). Parole violations and revocations in California: Analysis and suggestions for action. *Federal Probation*, *73*(1), 2.

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.

Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications* (2nd ed.). SAGE Publications, Inc.

Hamilton, Z. K. (2011). Adapting to bad news: Lessons from the Harlem Parole Reentry Court. *Journal of Offender Rehabilitation*, *50*(7), 385–410.

Hamilton, Z. K., & Campbell, C. M. (2013). A dark figure of corrections failure by way of participation. *Criminal Justice and Behavior*, *40*(2), 180–202.

Hamilton, Z. K., & Campbell, C. M. (2014). Uncommonly observed the impact of New Jersey's halfway house system. *Criminal Justice and Behavior*, *41*(11), 1354–1375.

Hamilton, Z. K., Campbell, C. M., van Wormer, J., Kigerl, A., & Posey, B. (2016). The impact of swift and certain sanctions: An evaluation of Washington State's policy for offenders on community supervision. *Criminology & Public Policy*, *15*(4), 1009–1072.

Hamilton, Z., Kigerl, A., Campagna, M., Barnoski, R., Lee, S., van Wormer, J., & Block, L. (2016). Designed to fit: The development and validation of the STRONG-R ecidivism risk assessment. *Criminal Justice and Behavior*, *43*(2), 230–263.

Hamilton, Z., Kigerl, A., & Hays, Z. (2015). Removing release impediments and reducing correctional costs: Evaluation of Washington State's Housing Voucher Program. *Justice Quarterly*, *32*(2), 255–287.

Harrell, A., & Roman, J. (2001). Reducing drug use and crime among offenders: The impact of graduated sanctions. *Journal of Drug Issues*, *31*(1), 207–231.

Harrison, P. (2017). Phone interview and review of this manuscript with Paige Harrison, former Director of Research at Washington State Department of Corrections.

Hawken, A. (2010). Behavioral triage: A new model for identifying and treating substance-abusing offenders. *Journal of Drug Policy Analysis*, *3*(1), 1–86.

Hawken, A., Kulick, J., Smith, K., Mei, J., Zhang, Y., Jarman, S., Yu, T., Carson, C., & Vial, T. (2016). *HOPE II: A follow-up to Hawai'i's HOPE evaluation* (No. NIJ 249912). School of Public Policy, Pepperdine University.

Holtfreter, K., & Wattanaporn, K.A. (2014). The transition from prison to community initiative: An examination of gender responsiveness for female offender reentry. *Criminal Justice and Behavior*, *41*(1), 41–57.

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data*. Wiley-Interscience.

Jolliffe, D., & Hedderman, C. (2015). Investigating the impact of custody on reoffending using propensity score matching. *Crime & Delinquency*, *61*(8), 1051–1077.

Kleiman, M. A. R., Kilmer, B., & Fisher, D. T. (2014). Theory and evidence on the swift-certain-fair approach to enforcing conditions of community supervision. *Federal Probation*, *78*, 71.

Kubrin, C. E., & Stewart, E. A. (2006). Predicting who reoffends: The neglected role of neighborhood context in recidivism studies. *Criminology*, *44*(1), 165–197.

Kurth, T., Walker, A. M., Glynn, R. J., Chan, K. A., Gaziano, J. M., Berger, K., & Robins, J. M. (2006). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *American Journal of Epidemiology*, *163*(3), 262–270.

Labrecque, R. M., Smith, P., Lovins, B., & Latessa, E. J. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation*, *53*(2), 116–128.

Latimore, P. K., MacKenzie, D. L., Dawes, D., & Tueller, S. (2016). Outcome findings from the HOPE demonstration field experiment: Is swift, certain, and fair an effective supervision strategy? *Criminology & Public Policy*, *15*(4), 1103–1141.

Lee, W.-S. (2011). Propensity score matching and variations on the balancing test. *Empirical Economics*, *44*(1), 47–80.

Leite, W. (2017). *Practical propensity score methods using R*. SAGE Publications.

Lipsey, M., Landenberger, N. A., Wilson, S. J., Lipsey, M., Landenberger, N. A., & Wilson, S. J. (2007). Effects of cognitive-behavioral programs for criminal offenders: A systematic review. *Campbell Systematic Reviews*, *3*(6), 1–27.

Lowenkamp, C. T., Flores, A. W., Holsinger, A. M., Makarios, M. D., & Latessa, E. J. (2010). Intensive supervision programs: Does program philosophy and the principles of effective intervention matter? *Journal of Criminal Justice*, *38*(4), 368–375.

Lowenkamp, C. T., & Latessa, E. J. (2005). Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology & Public Policy*, *4*(2), 263–290.

Lutze, F. E., Rosky, J. W., & Hamilton, Z. K. (2014). Homelessness and reentry a multisite outcome evaluation of Washington State's reentry housing program for high risk offenders. *Criminal Justice and Behavior*, *41*(4), 471–491. https://doi.org/10.1177/0093854813510164

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

Morash, M., Kashy, D. A., Smith, S. W., & Cobbina, J. E. (2015). The effect of probation or parole agent relationship style and women offenders' criminogenic need on offenders' responses to supervision interactions. *Criminal Justice and Behavior*, *42*(4), 412–434.

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.

Morash, M., Kashy, D. A., Smith, S. W., & Cobbina, J. E. (2019). Technical violations, treatment and punishment responses, and recidivism of women on probation and parole. *Criminal Justice Policy Review*, *30*(5), 788–810.

O'Connell, D. J., Brent, J. J., & Visher, C. A. (2016). Decide your time: A randomized trial of a drug testing and graduated sanctions program for probationers. *Criminology & Public Policy*, *15*(4), 1073–1102.

Onifade, E., Wilkins, J., Davidson, W., Campbell, C., & Petersen, J. (2011). A comparative analysis of recidivism with propensity score matching of informal and formal juvenile probationers. *Journal of Offender Rehabilitation*, *50*(8), 531–546.

Ostermann, M. (2009). An analysis of New Jersey's day reporting center and halfway back programs: Embracing the rehabilitative ideal through evidence based practices. *Journal of Offender Rehabilitation*, *48*(2), 139–153.

Ostermann, M., & Salerno, L. M. (2016). The validity of the level of service inventory – Revised at the intersection of race and gender. *The Prison Journal*, *96*(4), 554–575.

Padgett, K. G., Bales, W. D., & Blomberg, T. G. (2006). Under surveillance: An empirical test of the effectiveness and consequences of electronic monitoring. *Criminology & Public Policy*, *5*(1), 61–91.

Petersilia, J., & Turner, S. (1993). Intensive probation and parole. *Crime and Justice*, *17*, 281–335.

Prendergast, M. L., Pearson, F. S., Podus, D., Hamilton, Z. K., & Greenwell, L. (2013). The Andrews' principles of risk, needs, and responsivity as applied in drug treatment programs: meta-analysis of crime and drug use outcomes. *Journal of Experimental Criminology*, *9*(3), 275–300.

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*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Rettinger, L. J., & Andrews, D. A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior*, *37*(1), 29–46.

Rosenbaum, P. R. (2002). *Observational studies*. Springer.

Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. In *Encyclopedia of statistics in behavioral science* (Vol. 4, pp. 1809–1814). John Wiley & Sons, Ltd.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38.

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.

Salisbury, E. J., Van Voorhis, P., & Spiropoulos, G. V. (2009). The predictive validity of a gender-responsive needs assessment: An exploratory study. *Crime & Delinquency*, *55*(4), 550–585.

Saxena, P., Messina, N. P., & Grella, C. E. (2014). 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.

Skeem, J., Monahan, J., & Lowenkamp, C. (2016). Gender, risk assessment, and sanctioning: The cost of treating women like men. *Law and Human Behavior*, *40*(5), 580–593.

Shaffer, D. K. (2011). Looking inside the black box of drug courts: A meta-analytic review. *Justice Quarterly*, *28*(3), 493–521.

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy*, *8*(1), 183–208.

Spence, D. H., & Haas, S. M. (2015). Predicting client success in day report centers: The importance of risk and needs assessment. *Journal of Offender Rehabilitation*, *54*(7), 502–519.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1–21.

Taxman, F. S. (1999). Unraveling "what works" for offenders in substance abuse treatment services. *National Drug Court Institute Review*, *2*(2), 93.

Taxman, F. S. (2012). Crime control in the twenty-first century: Science-Based Supervision (SBS). *Journal of Crime and Justice*, *35*(2), 135–144.

Taxman, F. S., & Belenko, S. (2012). Current state of EBP in community corrections field. In *Implementing evidence-based practices in community corrections and addiction treatment* (pp. 151–188). Springer.

Taxman, F. S., Byrne, J. M., & Young, D. W. (2002). *Targeting for reentry: Matching needs and services to maximize public safety*. Bureau of Governmental Research, University of Maryland.

Taxman, F. S., Soule, D., & Gelb, A. (1999). Graduated sanctions: Stepping into accountable systems and offenders. *The Prison Journal*, *79*(2), 182–204.

van Wormer, K., & Kaplan, L. E. (2006). Results of a national survey of wardens in women's prisons: The case for gender specific treatment. *Women & Therapy*, *29*(1/2), 133–151.

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.

Vito, G. F., Higgins, G. E., & Tewksbury, R. (2015). The effectiveness of parole supervision use of propensity score matching to analyze reincarceration rates in Kentucky. *Criminal Justice Policy Review*, *28*(7), 627–640.

Washington State Department of Corrections. (2014). *Gender Responsiveness*, DOC 590.370. Prison/Work Release/Field. www.doc.wa.gov/information/policies/files/590370.pdf

Wattanaporn, K. A., & Holtfreter, K. (2014). The impact of feminist pathways research on gender-responsive policy and practice. *Feminist Criminology*, *9*(3), 191–207.

White, M. D., Mellow, J., Englander, K., & Ruffinengo, M. (2011). Halfway back: An alternative to revocation for technical parole violators. *Criminal Justice Policy Review*, *22*(2), 140–166.

Wodahl, E. J., Boman IV, J. H., & Garland, B. E. (2015). Responding to probation and parole violations: Are jail sanctions more effective than community-based graduated sanctions? *Journal of Criminal Justice*, *43*(3), 242–250.

Wodahl, E. J., Ogle, R., Kadleck, C., & Gerow, K. (2013). Offender perceptions of graduated sanctions. *Crime & Delinquency*, *59*(8), 1185–1210.

Wojtowicz, J. P., & Tongyin, L. (2006). A seven-year analysis of walk-away rates of the New Jersey Halfway House Program. *Corrections Compendium*, *31*(6), 6–23.

Wooldridge, J. M. (2005). Violating ignorability of treatment by controlling for too many factors. *Econometric Theory*, *21*(05), 1026–1028.

Yesberg, J. A., Scanlan, J. M., Hanby, L. J., Serin, R. C., & Polaschek, D. L. L. (2015). Predicting women's recidivism: Validating a dynamic community-based 'gender-neutral' tool. Probation Journal, *62*(1), 33–48.

Zhang, S. X., Roberts, R. E. L., & Callanan, V. J. (2006). The cost benefits of providing community-based correctional services: An evaluation of a statewide parole program in California. *Journal of Criminal Justice*, *34*(4), 341–350.

Author Biographies

Christopher M. Campbell, PhD, is an assistant professor in the Department of Criminology and Criminal Justice at Portland State University. His research interests include institutional and community corrections, offender rehabilitation and reentry, juvenile justice, and the application of criminological theory. His recent work has appeared in the *Journal of Criminal Justice*, *Criminal Justice and Behavior*, *Criminology and Public Policy*, and *Justice Quarterly*.

Ryan M. Labrecque, PhD, is an assistant professor in the Department of Criminal Justice at the University of Central Florida. His research focuses on the evaluation of

correctional interventions, the effects of prison life, the development of risk and needs assessments for community and institutional corrections settings, and the transfer of knowledge to practitioners and policy makers. His work has appeared in *Crime and Delinquency; Criminal Justice and Behavior; Journal of Criminal Justice; Psychology*, *Public Policy*, and *Law*; and *Violence and Victims*.

Megan E. Mohler, MS, is a doctoral student at Temple University and works as a freelance research analyst. She earned her master's from the Department of Criminology and Criminal Justice at Portland State University. Her research interests include gender differences in crime and criminal justice, and the interaction of psychological and sociological influences in police and community interactions. Her thesis examined the way social media influences the role of public perceptions of procedural justice toward the police.

Molly J. Christmann, MS, is a managment analyst with the City of Portland's Independent Police Review. She earned her master's from the Department of Criminology and Criminal Justice at Portland State University. Her research interests include gender differences in crime and criminal justice, implementation of criminal justice practices, crime analysis, and geographical information systems. Her work has recently appeared in *Corrections: Policy*, *Practice and Research*.