Published in final edited form as:

J Subst Abuse Treat. 2014 March ; 46(3): 332–339. doi:10.1016/j.jsat.2013.09.005.

Gender-Sensitive Substance Abuse Treatment and Arrest Outcomes for Women

Wendy B. Kissin,

Westat, 1600 Research Boulevard, Rockville, Maryland 20850, USA

Zhiqun Tang,

Westat, 1600 Research Boulevard, Rockville, Maryland 20850, USA

Kevin M. Campbell,

Division of Behavioral Health and Recovery, Washington State Department of Social and Health Services, Box 45330, Olympia, Washington 98504-5330.

Ronald E. Claus, and

Westat, 1600 Research Boulevard, Rockville, Maryland 20850, USA

Robert G. Orwin

Westat, 1600 Research Boulevard, Rockville, Maryland 20850, USA

Abstract

The present study links an empirically-developed quantitative measure of gender-sensitive (GS) substance abuse treatment to arrest outcomes among 5,109 substance abusing women in mixed-gender short-term residential programs in Washington State. Frailty models of survival analysis and 3-level hierarchical linear models were conducted to test the beneficial effects of GS treatment on decreasing criminal justice involvement. Propensity scores were used to control for the pre-existing differences among women due to the quasi-experimental nature of the study. Men's arrest outcomes were used to control for confounding at the program level. Results show that women in more GS treatment programs had a lower risk of drug-related arrests, and women in more GS treatment programs who also completed treatment had a significant reduction in overall arrests from two years before- to two years after treatment, above and beyond the reduction in arrests due to treatment alone. Implications and directions for future research are discussed.

Keywords

gender-sensitive treatment; drug abuse treatment; women offenders; arrest; criminal justice outcomes

1. Introduction

In recent years, the United States has seen a dramatic increase in the number of women involved in the criminal justice (CJ) system. From the year 2000 to 2010, the number of

© 2013 Elsevier Inc. All rights reserved.

Correspondence concerning this article should be addressed to Wendy Kissin, Westat, 1600 Research Boulevard, Rockville, MD 20850, United States. wendykissin@westat.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

women arrested increased by 29% while the number of men arrested increased by only 7% (Bureau of Justice Statistics [BJS], 2011a). Although women accounted for fewer than 7% of individuals under State or Federal prison jurisdiction in 2010, this represents a 23% increase since the year 2000 and 159% increase since 1990 (BJS, 2011b). Increased incarceration of women has been attributed to mandatory sentencing for drug offenses (Austin, Bruce, Carroll, McCall, & Richards, 2001; Mauer, Potler, & Wolf, 1999). Upwards of 80% of female offenders have a substance abuse problem, and 50% of women are under the influence of substances at the time they commit the offense (Bloom, Owen & Covington, 2003; Covington & Bloom, 2006).

Substance abuse is a primary risk factor for involvement in the CJ system. Substance abuse is one of the “central eight” risk and/or need factors for CJ involvement for both women and men (Andrews, Bonta, & Wormith, 2006). Other risk factors include problems with marital/family relationships, education/employment, and leisure activities, and the “Big 4”, or personality characteristics, criminal thinking, criminal history, and criminally-involved peers (Andrews et al., 2006). However, the development and manifestation of substance abuse among women can be more severe than for men, involving more physical, psychological, family/social, and socioeconomic problems (Messina, Calhoun & Warda, 2012; Pelissier & Jones, 2005; Spjeldnes & Goodkind, 2009). Cumulative evidence about risks for CJ involvement specific to women comes from recidivism studies among women offenders primarily during the past decade. The most common reasons for recidivism among women offenders include substance abuse and mental health problems; a history of being physically and/or sexually abused (Messina, Grella, Burdon, & Prendergast, 2007; McDonald, 2008; Salisbury & Van Voorhis, 2009); and parenting stress with minimal economic and emotional support for child rearing (Van Voorhis et al., 2010). Salisbury and Van Voorhis (2009) identified three common pathways to offending among women: childhood victimization, unhealthy relationships, or (the more gender-neutral) financial problems. The former two pathways, childhood victimization and dysfunctional relationships, contribute to the development of depression, anxiety and other mental health problems, substance abuse, and eventual involvement in the CJ system.

Effective treatment for women's substance use disorders can reduce their involvement with the CJ system (Mosher & Phillips, 2006). Prison-based substance abuse treatment has been shown to reduce rates of reoffending among women (Dowden & Blanchette, 2002; Mosher & Phillips, 2006; Pelissier, Camp, Gaes, Saylor, & Rhodes, 2003; Pelissier, Motivans, & Rounds-Bryant, 2005; Robbins, Martin, & Surratt, 2009), especially among women who complete treatment (Mosher & Phillips, 2006; Robbins et al., 2009) or spend more time in treatment (Messina, Burdon, Hagopian, & Prendergast, 2006). Moreover, in response to the multidimensional profile of problems women display upon admission to substance abuse treatment and correctional settings, emerging literature shows that researchers in both fields have been questioning the traditional policies and practices developed largely “through the lens of managing men, not women” (Ney, Ramirez, & Van Diemen, 2012, p.4). Researchers and practitioners in the substance abuse and the CJ fields emphasize the importance of understanding gender differences and women's specific needs and risk factors, and advocate for comprehensive, strengths-based, relational, trauma-informed, and gender-sensitive (GS) (or gender-responsive) treatment provided in a safe, affirming environment (Bloom et al., 2003; Covington & Bloom, 2006; CSAT, 2009; Greenfield et al., 2007; Greenfield & Grella, 2009; Grella, 2008; Ney et al., 2012; Van Voorhis et al., 2010).

GS treatment has been effective in correctional settings (Van Voorhis, 2012), and has led to decreased drug use, criminal behavior, and trauma exposure and increased mental health time before reincarceration (Sacks, McKendrick, & Hamilton, 2012). Several prison-based GS substance abuse treatment programs, such as the California Institute of Women's Forever

Free Program, Helping Women Recover (Covington, 2008) and Beyond Trauma (Covington, 2003), have been developed to address the needs of women offenders. Indeed, women offenders who participate in prison-based GS substance abuse treatment programs have been more likely to remain drug-free, to have lower CJ recidivism rates than the comparison group (Messina, Grella, Cartier, & Torres, 2010; Messina et al., 2012; Wells & Bright, 2005), and to spend more time engaged in aftercare (Messina et al., 2010).

Most studies in this area focus on substance abuse treatment and CJ outcomes for women offenders during and/or following probation, incarceration, or parole (i.e., a 100% offending sample). Relatively few studies have examined women's CJ outcomes following treatment among general substance abuse treatment admissions, and even fewer have assessed the effects of GS treatment on CJ outcomes among a general substance abuse treatment population of women (Ashley, Marsden, & Brady, 2003). Some evidence for the relationship between GS treatment and favorable CJ outcomes comes from a limited number of studies comparing women's outcomes from women-only to mixed-gender programs. In a study involving pregnant women (Daley et al., 2000), lower levels of criminal activity were reported following substance abuse treatment provided in multiple modalities (i.e., outpatient, methadone, residential, detox, residential/outpatient), with greater decreases noted among women with vocational skills and among women with more severe drug use. In this study, the cost savings of avoided crime more than paid for treatment costs (Daley et al., 2000). A more recent study (Prendergast, Messina, Hall, & Warda, 2011) reported decreased substance abuse, criminal activity, and arrests and increased employment among women 12 months after entering outpatient substance abuse treatment; in addition, women who attended women-only treatment showed greater decreases in substance abuse and self-reported criminal activity (but no difference in arrest rates) than women who attended mixed-gender treatment. Similarly, pregnant and parenting women who attended women-only (vs. mixed-gender) treatment had fewer arrests and better overall outcomes at 10-year followup (Evans, Li, Pierce, & Hser, 2013). In these studies, 61% of the women had been arrested in the year prior to treatment (Prendergast et al., 2011), 70% reported lifetime arrests, and 49% reported lifetime incarceration (Daley et al., 2000).

Most women who receive substance abuse treatment in the United States do so in mixed-gender settings (Grella & Greenwell, 2004), and there is considerable variation in the extent to which these programs are tailored to meet women's needs (Tang, Claus, Orwin, Kissin, & Arieira, 2012). One evaluation of the potential value added by GS substance abuse treatment is to examine women's post-treatment arrests. Based on a comprehensive review of the literature, Grella (2008) proposed a multidimensional model for measuring GS substance abuse treatment services and practices, which Tang and colleagues (2012) implemented in their development of a quantitative measure. This measure of GS treatment was developed using organization-level data collected from mixed-gender, intensive inpatient programs (IIPs, i.e., short-term residential programs) in Washington State, which it placed empirically along a continuum of GS treatment. In general, programs ranked as more GS tended to report offering the services, staff expertise, and milieu to provide GS treatment as described here. Theoretically, GS substance abuse treatment may directly or indirectly reduce women's CJ involvement in different ways. First, it may directly lead to the reduction of substance abuse, above and beyond treatment as usual. Second, GS treatment might have direct effects on factors that make women vulnerable to criminal behavior that lead to or are accompanied by substance use, such as trauma, mental health problems, parenting and relationship stress, and financial problems. Third, GS treatment may indirectly affect CJ outcomes by increasing women's engagement in care, which has been shown to improve treatment outcomes (Moos & Moos, 2003).

The present study uses administrative data to examine the effects of GS treatment on post-treatment arrests among women treated in mixed-gender substance abuse treatment programs. The primary hypothesis is that women treated in more GS programs will have fewer arrests overall and fewer drug-related arrests during the first two years post-treatment relative to women treated in less GS programs. The secondary hypothesis is that more GS treatment is associated with still fewer post-treatment arrests among women who complete treatment.

2. Materials and Methods

2.1 Study Sample

The study included 13 mixed-gender IIP programs in Washington State. Washington's Department of Social and Health Services regulations require that IIP programs provide at least 20 hours of treatment per week, including a minimum of 10 hours of chemical dependency counseling per week. These IIP programs offer short-term (21 or 30 days, depending on the program) residential substance abuse treatment to women and men age 18 and over and serve both public- and private-pay clients. Programs were free to decide what specific curriculum or treatment approach to use. Programs were eligible for inclusion in the study if they had a client gender ratio no smaller than 1:11 in either direction, thus affording appropriate variability in a mixed-gender environment. All eligible programs agreed to participate. During the study period from April 1, 2005 to March 31, 2009, from 277 to 2,332 publicly-funded women were admitted to each program. The percentage of women in each program varied widely, from 12% to 91%.

The client sample was comprised of 5,109 women and 9,838 men admitted to the programs during the study period. For inclusion in the study, clients had to be qualified for publicly-funded treatment (based on income at or below 185% of the federal poverty level). Only data from publicly-funded clients were included in the study. Human subjects approval was obtained.

Women admitted to the 13 mixed-gender IIPs had a mean age of 34.2 ($SD = 10.2$, $range = 18 - 77$) and education of 11.4 years. During the 30 days before treatment, 24.5% of the women were employed (based on receiving income from employer-reported employment). Relatively few participants (16%) were married or in a committed relationship, and 43% had never been married. The racial/ethnic composition was 76% Caucasian; 7% each of Native American, African American, and other races; and 3% multiracial with 8% reporting Hispanic ethnicity (measured independently of race). Substances used most commonly during the 30 days prior to treatment entry were alcohol (67%), methamphetamine (47%), marijuana (45%), and cocaine (31%). Heroin use and past 30-day needle use were relatively rare at 11% and 14%, respectively.

At treatment admission, 55.6% of women reported feeling moderately to extremely troubled by a mental health problem during the past 30 days (with 15.2% missing data), although mental health symptom and diagnosis data are unavailable. Many (60.8%) women reported a lifetime history of domestic violence, though data are missing for 9.8% of cases. In terms of treatment, 28.7% of women reported currently taking legally-prescribed psychiatric medication, 25.7% said they received mental health treatment during prior 12 months, and staff rated 34.2% of women as receiving or in need of mental health care.

2.2 Measures

This study relies on administrative data for all client-level measures. The GS measure was based on data collected during program site visits.

2.2.1 GS treatment—On site visits to each program, researchers conducted semi-structured interviews with a counselor, the clinical director, and the program director, then completed an observational protocol. Interview data included comprehensive information about the treatment programs, with special attention to GS aspects of the programs for women and men (Tang et al., 2012). Rasch modeling used site visit data to generate three continuous program-level scales measuring aspects of GS treatment: Treatment Orientation and Staff Training, Women's and General Services, and Physical Environment (Tang et al., 2012), which were based on domains identified in Grella's (2008) review of the GS treatment literature. Content of the 16-item Treatment Orientation and Staff Training scale focuses on the program's mission and support for training in the areas of behavioral health (e.g., psychiatric and trauma topics) and GS treatment for women (e.g., women's addiction, sexuality, parenting, family violence). The 25-item Women's and General Services scale measures assessments and services for women in general (e.g., parenting skills, trauma counseling), women's health (e.g., family planning, safer sex) and quality treatment (e.g., life skills and stage of change assessments). The 16-item Physical Environment scale measures privacy, safety, and livability of the program's facility and neighborhood. Scales are reliable and valid by Rasch standards (Linacre, 2003; Linacre, 2010; Wright, Linacre, & Gustafson, 1994). Each of the scales has excellent program reliability ($\alpha = .92 - .93$) and adequate to very good item reliability ($\alpha = .70 - .86$). Two experts reviewed the hierarchical structures of items in the final models for meaning and content validity. Once the continuous scales were clustered into ranked ordinal scales, program separation statistics demonstrated that each scale reliably discerns between the groups of programs. Data collection and development of these scales is described as the sole focus of another article (Tang et al., 2012).

The composite GS variable used as the independent variable in the current analyses was constructed using the three GS scales just described. The composite variable consists of four groups of programs ranked from low to high gender sensitivity (*range* = 0 - 3). Given the small sample size at the program level, we took this more conservative approach rather than running multiple tests using the three different scales. This 4-level ordinal variable was constructed from the three continuous GS scales by assigning a weight of two to the Women's and General Services scale and a weight of one to the Treatment Orientation and Staff Training scale and to the Physical Environment scale then summing the scores. This weighting scheme was chosen based on correlations between the individual scales (Tang et al., 2012) and on our clinical judgment that treatment services are more important to a program's overall gender sensitivity than the other two domains. With that being said, due to the small number of programs, the resulting continuous score was converted into an ordinal scale that ranks each program's gender sensitivity as 0 = low (3 programs), 1 = moderately low (3 programs), 2 = moderately high (4 programs) or 3 = high (3 programs) based on recommended methods (Orwin, Sonnefeld, Cordray, Pion, & Perl, 1998). The selection of ordinal cut points took advantage of the natural grouping of programs by Rasch scores and was also guided by the standard errors, leading to a sufficient proportion of women (15.3 - 32.2%) and a similar number of programs in each category.

2.2.2 Outcomes—Arrest data were provided by the Washington State Patrol. Original data were de-identified by the data supplier at Washington State's Department of Social and Health Services, and included the number of days per month (i.e., 30-day unit) a person was arrested for a felony or gross misdemeanor within the State and the type of offense over two years pre- and two years post- treatment. Audits verified the completeness of felony and gross misdemeanor records. Petty misdemeanors were excluded because audits were not available to verify the completeness of the data. In addition, exclusion of petty misdemeanors is intuitive because they are minor “non-criminal” offenses (e.g., traffic violations) carrying a maximum punishment of a few-hundred dollar fine.

For survival analysis, the two outcome measures created were (a) days from treatment discharge to first arrest of any kind and (b) days from treatment discharge to first drug-related arrest. Arrests (any and drug-related) were relatively rare events. For hierarchical linear modeling (HLM), original monthly arrest data were collapsed into quarterly variables and then dichotomized (i.e., arrested vs. not arrested), thus representing a woman's arrest status every 90 days over the two years pre- and the two years post-treatment.

2.2.3 Propensity score—Due to the nonexperimental study design, the attribution of observed differences in client outcomes to program differences in GS treatment represents a significant challenge. Propensity scores (Rosenbaum & Rubin, 1983) were created to adjust baseline differences in women across the gender sensitivity groups. Bias related to treatment assignment was also mitigated by the construction of the GS treatment variable after “treatment assignment” occurred naturally (i.e., in real world treatment settings).

We selected covariates for the propensity score model using a standard approach (Rosenbaum & Rubin, 1983) but recognizing recent cautionary guidance (Pearl, 2010). Specifically, propensity score precision may be enhanced by including 1) potential confounders of the treatment-outcome relationship and 2) variables potentially associated with the study's treatment outcomes, as determined by past research or expert knowledge (Austin, Grootendorst & Anderson, 2007; Austin, 2008). Stepwise selection approaches (Guo & Fraser, 2010) were used to identify additional predictors of exposure to GS treatment.

Balance, or the similar multivariate distribution of the covariates across the treatment groups, was tested by examining effect-size differences of the model's terms in propensity score quintiles rather than relying solely upon statistical tests (and the resulting *p* values) that are subjected to the impact of sample size (Austin, 2008; Claus et al., 2007). A model is considered balanced when the distribution of covariates is similar within each quintile and the number of differences are no greater than one might expect to occur by chance (Rubin, 1997).

The propensity score model was developed using a series of ordinal logistic regression models, fit with a broad range of covariates or confounders in the areas of women client demographics, employment, income, living arrangements, mental health, substance use, CJ, and disability which were available from the intake data from the Treatment and Assessment Report Generation Tool (TARGET), the Washington State management information system, based on client self-report and interviewer ratings from the admission process. Using an iterative process, covariate interaction and polynomial terms were added to achieve balance. For the final propensity scores generated, 95.7% of the balance tests showed no larger than small effect size differences between gender sensitivity groups within the propensity quintiles, i.e., no more than would be expected by chance.

2.2.4 Men's outcomes—Confounding effects at the program-level present an additional challenge. Whatever led higher GS programs to incorporate such programming for women may have caused additional program improvements, reflect greater adaptability, or simply be a proxy for better management. Since these factors are not readily measurable, we used men's arrests of any kind and drug-related arrest outcomes to control for program-level confounders. Within evaluation methodology, this is a variation on the nonequivalent dependent variables design (Shadish, Cook & Campbell, 2002). It is nonequivalent in the sense that from a construct-validity-of-the-cause standpoint, the comparison sample (men) should not be affected by the part of the intervention that is GS, even though they might be affected by other program-level factors associated with it (e.g., program quality). This controls for the very plausible threat that more GS programs produce better outcomes for

women because they are more effective programs, independent of the level of GS treatment. The corresponding aggregated program-level outcomes for the 9,838 men who attended the 13 IIP programs during the study period were entered into models as a time-varying covariate. In survival analyses, men's arrests were controlled by using time-varying computed hazard rates of the first arrest and the first drug-related arrest among men treated in the same treatment program as women. In HLM, a time-varying program-level variable represented the proportion of men with any arrests or drug-related arrests during a particular interval (i.e., quarter).

2.2.5 Controlled environment—We controlled for times during which women were in a controlled environment because during these times (for the most part) they did not have the opportunity to commit offenses for which they could be arrested. The controlled environment variable included time spent in a residential or inpatient substance abuse, mental health, and/or medical setting as well as time spent in prison. The variable was coded yes/no and included as a daily time-varying covariate in survival analysis, and was coded as a continuous variable representing the number of days in each quarter spent in a controlled environment over 16 quarters.

2.2.6 Treatment completion—Treatment completion was based on the treatment program's indication that the client had completed treatment as reported in TARGET.

2.2.7 Client demographic characteristics—Client substance use and most demographic information at intake was obtained from the TARGET intake interview, as mentioned relative to the propensity score. Additional employment data were provided by the Washington State Employment Security Department and de-identified by the data supplier at Washington State's Department of Social and Health Services. Employment data were provided in the form of monthly personal income earned by legal means and reported to the State by employers.

2.3 Data Analyses

Survival analysis and HLM were used to examine the two hypotheses in this study. The two methods can complement each other, and used together can depict a more complete picture of the two arrest outcomes spanning two-year pre- and post- treatment. Specifically, HLM models all arrests across the entire four years whereas survival analysis focuses on the time to the first post-treatment arrest. In addition, due to small sample size at the program level, we hoped results would be congruent for both methods.

2.3.1 Survival analysis—To examine the first arrest of any kind and the first drug-related arrest during the two years after treatment discharge (in days), survival analysis using Cox proportional hazard model was performed with SAS version 9.3 (Cary, NC, USA). Survival analysis accommodates clients' different discharge and follow-up dates (and deaths, if relevant), and correlations among clients' data clustered within the same treatment program can be adjusted using frailty models (equivalent to a 2-level model with clients nested within programs). Women clients who did not have an arrest (or drug-related arrest for the corresponding outcome) were censored on the 730th day after discharge or upon death if it occurred before the 730th day without being preceded by an arrest.

Several models were performed to predict each of the two arrest outcomes, including a survival model without vs. with the frailty term (i.e., random effect), using the GS treatment as the key predictor. To rule out other plausible explanations (i.e., potential confounders), we controlled for women clients' propensity score, two-year pre-treatment arrest status, controlled environment during the two years post-treatment (yes/no day-by-day treated as a

time-varying covariate), and the aggregated program-level probability of men's arrests during the same quarter. If the frailty term was found significant, the frailty model was used in next steps and treatment completion was further included.

2.3.2 Hierarchical linear modeling (HLM)—HLM, version 7.0 was used to examine the longitudinal occurrence of arrests spanning two years pre- and two years post-treatment. Compared to survival analysis, HLM can model all repeated arrests with more than 2-levels of clustering. Three-level models were created to study quarterly arrest measurements nested within clients and clients nested within programs. Further, time variables were created to jointly model an outcome over two years pre- and two years post-treatment.

The three-level models in this study were specified using four growth variables representing a discontinuous piecewise linear growth model in Level-1 that had two separate intercepts and two linear slopes¹. The pre-treatment intercept represents the client's probability of arrests during the quarter preceding treatment admission, and the pre-treatment linear slope represents the quarterly change in probability of arrests during the two years prior to treatment. Similarly, the post-treatment intercept represents the client's probability of arrests during the first quarter after treatment admission, and the post-treatment linear slope represents the quarterly change in probability of arrests post-treatment. In short, the Level-1 model jointly estimated the pre-treatment and post-treatment individual arrest trajectories on the logit scale of the binary arrest outcome, both in a linear shape but with different intercepts and slopes.

The Level-1 model also controlled for women's days in a controlled environment as a quarterly time-varying covariate and the proportion of men from the same program who were arrested, also as a quarterly time-varying covariate. The propensity score was entered in the Level-2 model, to control for differences in client characteristics in the prediction of each of the four growth parameters. Each of the four growth parameters had a random component whereby the intercepts and slopes of the arrest outcome could vary across women clients. Further, the intercepts and slopes were allowed to vary across programs, leading to a total of four Level-3 random components. The GS treatment variable was entered in Level-3 to predict program-level intercepts and slopes. Thus, a GS effect was expressed by a significant effect of GS treatment on the intercept and/or the slope of the arrests during post-treatment. Pre-existing differences across GS levels were controlled for by using the GS variable predicting the intercept and the slope of the arrests during pre-treatment. However, those effects, if any, were anticipated non-significant.

Six steps were involved in HLM analyses. First, the base model was fit to the data, including the two time-varying covariates each as a fixed effect in Level-1, the propensity score as a fixed effect predicting the four growth parameters in Level-2, and four random components in each of Levels 2 and 3. The purpose of the first step was to examine the random components of the four growth parameters. Second, non-significant random components were dropped from the base model one by one and in a different order. The purpose of this step was to identify the final base model to ensure it only contained significant random components of the growth parameters. Although any non-significant random components of the growth parameters were dropped in this step, the two time-varying covariates in Level-1 and the propensity score in the four Level-2 equations were retained regardless of their significance. Third, the GS variable was entered to predict all four growth parameters in Level-3 four equations simultaneously, and then the non-significant GS variable was dropped from its corresponding Level-3 equation, one at a time, starting with the variable

¹Linear models were fit to the data because fitting models with quadratic slopes to the data did not converge or were otherwise problematic.

with the largest p-value. This step examined the relative importance of the GS variable in predicting growth parameters. Fourth, the GS variable was entered to predict each of the four growth parameters in Level-3 one at a time. Through Steps 3 and 4 it was ensured that any GS variable that was dropped from further analyses was indeed non-significant (by running the model backwards, forwards and in various orders). Fifth, treatment completion was entered as a fixed effect in Level-2 to predict post-treatment growth parameters and treatment completion was dropped from the corresponding equation when non-significant. Finally, a cross-level interaction was specified to use the GS variable in Level-3 to predict treatment completion in Level-2 for the corresponding growth parameter. The purposes of steps 5 and 6 were to examine whether treatment completion had any unique effect on post-treatment growth parameters, whether the GS variable remained predictive after adding treatment completion, and whether the GS variable was related to arrest outcomes for treatment completers if not for all women clients.

3. Results

Based on intake data, 8.5% of the women ($n = 433$) were court-mandated to receive substance abuse treatment. At treatment intake, over half of the women (51.6%) reported the following types of legal involvement during the prior month (clients could select multiple items): on probation or parole (25.7%), awaiting trial (7.4%), awaiting charges (6.0%), in adult drug court (5.6%), in other program supervised by a CJ-related agency (2.2%), and driving under the influence (DUI) deferred (1.4%). Fewer than 1% of women reported the following: in juvenile drug court, incarcerated pre-trial, on trial, convicted awaiting sentencing, incarcerated post-conviction, petitioning for DUI deferred prosecution, and diversion. (Child custody issues were not included because it could not be determined whether these involved legal charges brought against the client.)

Women's average rate of treatment completion was 77.2% ($SD = 8.8\%$) and ranged from 59.0% to 82.9% across programs. Women's mean length of stay was 21.7 ($SD = 10.4$) days, including clients who dropped out. Median stays were 28.0 days for completers and 8.0 days for noncompleters.

3.1 Survival analyses: GS effect on the first arrest during the two years post-treatment

3.1.1 Overall arrests—GS treatment significantly reduced the risk of arrest by 11.4% (hazard ratio [HR] = 0.886, $p = .023$) among woman clients who completed treatment. The final model was a frailty model with a significant interaction between the GS variable and treatment completion, controlling for women clients' propensity score, pre-treatment arrest status, controlled environment during the two years post-treatment (a daily time-varying covariate), and risk (or probability) of arrest among men (see Table 1). Although treatment completion was significant before adding the GS interaction term ($HR = 0.727$, $p < .001$), it no longer was in the final model. Forty-six percent of women (2,347 of 5,109) were arrested during the two years prior to treatment admission. Thirty-two percent of women (1,636 of 5,109) were arrested during the two years post-treatment.

All potential confounding variables in the model except the propensity score were statistically significant. The risk of the first arrest post-treatment was 1.6 times higher for women clients who were offenders during the two years pre-treatment. Spending any day in a controlled environment post-treatment greatly reduced the risk, by 47.5% ($HR = 0.525$, $p < .001$). Women attending programs where male clients averaged more post-treatment arrests were also more likely to be arrested during a follow-up quarter. Specifically, a 1-unit increase in men's aggregated hazard rate was associated with a 2.2 times higher risk of arrest for women clients ($HR = 3.2$, $p < .001$).

The random effect of program in the survival frailty model was statistically significant (frailty estimate = 0.024, $X^2 = 24.04$, adjusted $df = 7.26$, $p = .0014$), suggesting an unobserved heterogeneity across programs in the risk of arrest. In other words, the risk of women clients experiencing the first arrest during the two years post-treatment varied across treatment programs beyond what was predicted by the programs' level of gender sensitivity.

3.1.2 Drug-related arrests—GS treatment had a marginal favorable effect on women's first incidence of drug-related arrests during the two years post-treatment, reducing the risk by 11.1% ($HR = 0.889$, $p = .099$). In addition, treatment completion was a significant negative predictor of arrests during the two years post-treatment ($HR = 0.711$, $p < .001$). That is, the risk of experiencing drug-related arrests was 28.9% lower for women clients who completed the treatment than for women who did not. The final model was a frailty model with the GS variable and treatment completion as main effects, controlling for women clients' propensity score, pre-treatment drug-related arrest status, controlled environment post-treatment, and men's risk of drug-related arrest (see Table 1). No interaction was found between GS level and treatment completion. Over a quarter of women (1,306 of 5,109) were arrested due to drug-related offenses during the two years prior to treatment admission. Over fifteen percent of women (792 of 5,109) experienced a drug-related arrest during the two years post-treatment.

All confounders in the model were significant except the propensity score. The risk of the first drug-related arrest post-treatment was 2.02 times higher for women clients who were offenders during pre-treatment period compared to those who were not. Being in a controlled environment during the two years post-treatment significantly reduced the risk of arrest by 63.5% ($HR = 0.365$, $p < .001$). Women attending programs where male clients averaged more post-treatment arrests were also more likely to be arrested during a follow-up quarter ($HR = 3.46$, $p < .001$). The random effect of program of the survival frailty model was significant (frailty estimate = 0.050, $X^2 = 22.80$, adjusted $df = 7.28$, $p = .002$).

3.2 HLM analyses: GS effect on arrests during the two years post-treatment

3.2.1 Overall arrests—The final 3-level base model showed that women clients' arrests increased before the index treatment (odds ratio [OR] = 1.21, $p < .001$), significantly decreased immediately after treatment admission ($OR = 0.43$, $p = .002$), and decreased further over the two-year follow-up period ($OR = 0.91$, $p < .001$) (see Table 2). Across the models, the GS variable did not predict the pre-treatment intercept or the pre-treatment slope. That is, as anticipated, pre-treatment arrest trajectories did not vary by the level of GS treatment. The GS variable did not predict the post-treatment slope but showed promising effects on the post-treatment intercept. In the final 3-level model, the GS variable, treatment completion, and the interaction between the two were entered to predict the post-treatment intercept. Both the post-treatment intercept ($OR = 0.54$, $p = .0022$) and the slope ($OR = 0.91$, $p < .001$) of arrests remained negative and significant. Specifically, from pre- to post-treatment, clients' overall odds of arrest were reduced by 46% and further reduced by 9% over each post-treatment quarters after the first post-treatment quarter. The interaction between the GS variable and treatment completion was significant, showing that GS treatment had a favorable cross-level effect on treatment completers. Specifically, the odds of being arrested post-treatment were 16% lower for treatment completers for every unit increase in GS level ($OR = 0.84$, $p = .011$). In addition, treatment completion reduced the odds of arrest, but this main effect was no longer significant ($OR = 0.81$, $p = 0.10$) after the inclusion of the GS interaction term.

These models controlled for propensity scores on the pre-treatment intercept and slope and propensity scores on the post-treatment intercept and slope. Models also controlled for

women's days in a controlled environment as a quarterly time-varying covariate, and proportion male clients in the program arrested during the same quarter. The propensity score was not significant in any of the four Level-2 equations. Being in a controlled environment during the two years post-treatment reduced odds of arrest by 2% for every day therein ($OR = 0.98, p < .001$). Men's arrest rate was unrelated to women's odds of arrest.

As indicated by the significant variance components ($p < .001$), the probability of any arrest over time varied across women by both level and change during the pre- and post-treatment periods. Women's average probability of arrest also varied across programs ($p < .001$ for both level-3 variance components of the pre-treatment and post-treatment intercepts; the level-3 variance components of the two slopes were non-significant).

3.2.2 Drug-related arrests—The final 3-level base model showed that women clients' drug-related arrests increased during the two years preceding the index treatment ($OR = 1.30, p < .001$), significantly decreased immediately after IIP admission ($OR = 0.44, p < .001$), and decreased further over time after treatment ($OR = 0.83, p < .001$). Similar to the overall arrest models, across models fit to drug-related arrests, the GS variable did not predict the pre-treatment intercept or slope. That is, as expected, pre-treatment drug-related arrest trajectories did not vary by the level of GS treatment. Although the GS variable did not predict the post-treatment slope it significantly predicted the post-treatment intercept. In the final 3-level model (see Table 2) after including the GS variable and treatment completion predicting the post-treatment intercept, the negative post-treatment intercept became only marginally significant but the negative post-treatment slope remained significant ($OR = 0.83, p < .001$), as in the final base model. Specifically, in the final model, the odds of drug-related arrests decreased 27% from pre- to post-treatment, although no longer significant ($OR = 0.73, p = .093$) compared to the final base model. The odds kept decreasing by 17% each post-treatment quarter after the first quarter.

Above and beyond these changes over time, GS treatment in Level-3 had a significant unique main effect for all women clients on the post-treatment intercept of drug-related arrests. Specifically, the post-treatment intercept was 16% lower for every unit increase in GS level ($OR = 0.84, p = .028$). Treatment completion also had a unique main effect. The odds of drug-related arrests were further reduced by 29% for treatment completers over non-completers ($OR = 0.71, p < .001$).

Being in a controlled environment during the two years post-treatment reduced odds of drug-related arrests by 3% for every day therein ($OR = 0.97, p < .001$). Men's arrest rate was negatively associated with women clients' odds of arrest ($OR = 0.97, p < .001$). Again, the propensity score was not significant in any of the four Level-2 equations.

As indicated by significant variance components ($p < .001$), the probability of drug-related arrests over time varied across women by both level and change during the pre- and post-treatment periods. The average probability of drug-related arrests also varied across programs ($p < .001$ for both level-3 variance components of the pre-treatment and post-treatment intercepts; the level-3 variance components of the two slopes were non-significant).

4. Discussion

Consistent with findings from prior studies (Dowden & Blanchette, 2002; Mosher & Phillips, 2006; Pelissier et al., 2003; Pelissier et al., 2005; Robbins et al., 2009), women were less likely to be arrested during the two years after substance abuse treatment than during the two years before regardless of whether the arrest was related to drugs or not. Also

consistent with the literature (e.g., Grella & Rodriguez, 2011; Mosher & Phillips, 2006; Robbins et al., 2009), treatment completers were less likely to be arrested post-treatment.

In support of the primary hypothesis, treatment that was more GS was associated with fewer drug-related arrests among women clients during the two years after treatment, compared with treatment that was less GS. The effects of GS treatment, although marginal in survival analyses, were consistently and clearly demonstrated in HLM, which uniquely predicted the reduction of drug-related arrests from pre- to post-treatment. That is, GS treatment effects remained over and above the reduction in drug-related arrests related to simply participating in IIP treatment. These effects were retained even after controlling for treatment completion, drug-related arrests during the two years pre-treatment, being in a controlled environment post-treatment, selection bias, and treatment quality confounders.

Treatment that was more GS was associated with lower overall risk of arrest post-treatment primarily among women who completed treatment, in support of the secondary hypothesis. This was demonstrated by interactions between GS level and treatment completion in the prediction of the reduction of any arrests in both survival and HLM analyses. As with drug-related arrests, GS treatment had an effect on the post-treatment probability of any arrest (i.e., intercept) but did not appear to be associated with the longitudinal change in arrests overall (i.e., linear slope) in HLM analyses, suggesting that GS treatment took effect on these outcomes immediately following IIP discharge. This was unlikely to occur by chance alone given that the 8-quarter pre-treatment arrest trajectories did not vary by GS level.

Findings in the present study demonstrate the benefits of GS treatment in mixed-gender programs on reducing rates of arrest overall and drug-related arrests, consistent with substance abuse treatment and CJ literature that advocates for the use of GS treatment for women in treatment and correctional settings (e.g., CSAT, 2009; Greenfield & Grella, 2009; Ney et al., 2012). By targeting the risk and/or protective factors common to both substance abuse and criminal behaviors, naturally-occurring GS treatment in mixed-gender substance abuse programs appeared to favorably affect arrest outcomes. The GS measure used in the present study includes GS treatment elements such as how inviting, livable, and safe the residential areas of the treatment programs are; staff training on women's substance use, mental health, and other women-specific topics; the use of strengths-based treatment plans; the provision of vocational needs and life skills assessments and trauma counseling; and the connection with social supports and a concrete post-treatment housing plan at treatment discharge (Tang et al. 2012).

The mechanism by which GS treatment can reduce women clients' odds of arrest merits further research attention. Recent studies of women offenders suggest that GS treatment targets women's mental health and/or trauma-related issues, which are thought to be mediators of women's criminal behavior (Lynch, DeHart, Belknap, & Green, 2012; Salisbury & Van Voorhis 2009). Previous limited studies showed that women-specific programs reduced psychiatric symptoms and promoted psychological well-being (Orwin, Francisco, & Bernichon, 2001), but studies detecting these factors as mediating effects are sparse. In their pilot study with a small sample of women offenders, Messina and colleagues (2012) found beneficial effects of GS treatment using Helping Women Recover and Beyond Trauma, compared to standard mixed-gender treatment, in reducing posttraumatic stress disorder (PTSD) symptomology, drug use, and rearrest rates and increasing psychological well-being. Baseline PTSD was the primary predictor of reduced rearrest rates (Messina et al., 2012). Likewise, more GS mixed-gender programs may be more successful in reducing mental health and/or trauma-related symptoms that make women vulnerable to criminal behavior (e.g., through poor affect regulation and poor decision-making in relationships and other parts of their lives (personal communication with Grella, March 2013)).

Although arrests data are limited to felonies and gross misdemeanors in Washington State, they are the most accurate arrest records in WA and are much more costly to society than petty misdemeanors. Providing GS treatment in mixed-gender substance abuse treatment programs can benefit society in terms of favorable CJ outcomes and reduced substance abuse among women. Such benefits are likely to outweigh the cost of treatment (Daley et al., 2000). The IIP programs in this study who implemented more GS treatment received the same funds as the other IIP programs, which suggests that choices made in allocating treatment funds are also important, not just the total level of funds. Findings from the present study are consistent with policies and practices of implementing GS treatment for women in substance abuse treatment (e.g., CSAT, 2009; Greenfield & Grella, 2009; Grella, 2008) and those whose substance abuse is treated in correctional settings (e.g., Messina et al., 2010; Messina et al., 2012; Ney et al., 2012).

This study has several methodological strengths. First, critical objective measures, including arrests, were based on administrative records data. Second, the study tested and used 3-level models in HLM to appropriately deal with the natural clustering of data spanning two years pre- and two years post-treatment. In survival analyses, frailty models were tested and used so that standard errors were adjusted for clustering. Third, data analyses used propensity scores to control for selection bias resulting from various preexisting differences across women, and further controlled for women's prior arrest history, days in controlled environment, and program confounders. Although theoretically there could still be selection bias due to unmeasured covariates, alternative explanations due to a broad variety of measured covariates at the client-level and at the program-level regarding GS treatment effects were largely reduced.

As is often a challenge in organizational research, the sample size at the program-level was small in this study. We could not use HLM robust test statistics that are appropriate for moderate to large sample sizes but not for the smaller sample size of programs employed here. Following well-designed, systematic and sequential modeling analytical procedures, all HLM models have shown results and effects consistently and it was unlikely that the study findings were simply observed by chance. Moreover, findings from two statistical methods both converged to largely support the hypothesized relationship between GS treatment and post-treatment arrests.

One next step would be to test the generalizability of these findings by applying the GS scale(s) to IIP programs in other areas, with other subgroups of women substance abusers, and other types of mixed-gender substance abuse programs. In this study, data were limited to publicly-funded clients treated in mixed-gender IIP programs in Washington State, which offers relatively comprehensive coverage to those in need. Most women were Caucasian, few were employed, almost half used methamphetamine, and relatively few used opioids. The GS scales themselves are likely to generalize to other program samples. Though we observed substantial variability in the program distribution on the three GS scales, the specific ordinal categorization of programs will not necessarily generalize because other samples of programs will vary in their rankings on the scales.

In conclusion, the present study addressed the value added by GS treatment to arrest outcomes among substance abusing women in mixed-gender IIPs. Women clients in more GS treatment programs had a lower risk of drug-related arrests after treatment, above and beyond the effect of treatment completion. Both the gender sensitivity of the program and treatment completion explain the reduction in drug-related arrests from pre- to post-treatment. For overall arrests, GS treatment reduced the risk among women clients who completed the treatment. This study further advances the quantitative organization-level measure of naturally-occurring GS substance abuse treatment services and practices for

women in mixed-gender programs (Tang et al., 2012) by using the composite GS measure to predict arrest outcomes among substance abusing women. The study begins to fill the knowledge gap about GS treatment as it naturally occurs in mixed-gender short-term residential programs. It also supports the effectiveness of GS treatment in improving key treatment outcomes for women such as CJ involvement.

Acknowledgments

The authors wish to thank NIDA for its financial support of this study. Sincere thanks go to Dr. Carlos Arieira (formerly at Westat) for data management and database preparation, to Dr. Alice Huber and Sue Green at Washington State's Department of Social and Health Services, Division of Behavioral Health and Recovery for their administrative, data support, and consultation regarding program operations, to Dr. Christine Grella at UCLA for her expert consultation on gender-sensitive treatment and the contents of this manuscript, to Allison Schlosser for her assistance with data collection and data entry, and to Dr. Shelly Greenfield at McLean Hospital/Harvard Medical School for her thoughtful and constructive review of the interview protocols.

This research was supported by a grant from the National Institute on Drug Abuse (NIDA R01DA020082-02), awarded to Robert G. Orwin, Ph.D.

References

- Andrews DA, Bonta J, Wormith JS. The recent past and near future of risk and/or need assessment. *Crime & Delinquency*. 2006; 52(7):7–27.
- Ashley OS, Marsden ME, Brady TM. Effectiveness of substance abuse treatment programming for women: A review. *The American Journal of Drug and Alcohol Abuse*. 2003; 29(1):19–53. [PubMed: 12731680]
- Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine*. 2008; 27:2037–2049. [PubMed: 18038446]
- Austin J, Bruce MA, Carroll L, McCall PL, Richards SC. The use of incarceration in the United States. *Critical Criminology*. 2001; 10:17–41.
- Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in Medicine*. 2007; 26(4):734–753. [PubMed: 16708349]
- Bloom, B.; Owen, BA.; Covington, S. *Gender-Responsive Strategies: Research, Practice, and Guiding Principles for Women Offenders*. US Dept. of Justice, National Institute of Corrections; 2003.
- Bureau of Justice Statistics. *Sourcebook of Criminal Justice Statistics, 2011, Section 4: Characteristics and Distribution of Persons Arrested*. U.S. Department of Justice; Washington, DC: 2011a.
- Bureau of Justice Statistics. *Sourcebook of Criminal Justice Statistics, 2011, Section 6: Persons Under Correctional Supervision*. U.S. Department of Justice; Washington, DC: 2011b.
- Center for Substance Abuse Treatment (CSAT). *Substance abuse treatment: Addressing the specific needs of women. Treatment Improvement Protocol (TIP) Series 51*. HHS Publication No. (SMA) 09-4426. Substance Abuse and Mental Health Services Administration; Rockville, MD: 2009.
- Claus RE, Orwin RG, Kissin W, Krupski A, Campbell K, Stark K. Does gender-specific substance abuse treatment for women promote continuity of care? *Journal of Substance Abuse Treatment*. 2007; 32:27–39. [PubMed: 17175396]
- Covington, SS. *Beyond Trauma Facilitator's Guide*. Hazelden Publishing; Center City, MN: 2003.
- Covington, SS. *Helping Women Recover: A Program for Treating Substance Abuse. Facilitator's Guide - Revised Edition for Use in the Criminal Justice System*. Jossey-Bass Publishers; San Francisco, CA: 2008.
- Covington SS, Bloom BE. Gender-responsive treatment and services in correctional settings. *Women and Therapy*, Elaine Leeder (Ed.). 2006; 29(3/4):9–33.
- Daley M, Argeriou M, McCarty D, Callahan JJ, Shepard DS, Williams CN. The costs of crime and the benefits of substance abuse treatment for pregnant women. *Journal of Substance Abuse Treatment*. 2000; 19:445–458. [PubMed: 11166509]

- Dowden C, Blanchette K. An evaluation of the effectiveness of substance abuse programming for female offenders. *International Journal of Offender Therapy and Comparative Criminology*. 2002; 24(2):220–230. [PubMed: 12113164]
- Evans E, Li L, Pierce J, Hser Y-I. Explaining long-term outcomes among drug-dependent mothers treated in women-only versus mixed-gender programs. *Journal of Substance Abuse Treatment*. 2013; 45:293–301. [PubMed: 23702103]
- Greenfield SF, Brooks AJ, Gordon SM, Green CA, Kropp F, McHugh RK, Lincoln M, Hien D, Miele GM. Substance abuse treatment entry, retention, and outcome in women: a review of the literature. *Drug and Alcohol Dependence*. 2007; 86:1–21. [PubMed: 16759822]
- Greenfield SF, Grella CE. What is “women-focused” treatment for substance use disorders? *Psychiatric Services*. 2009; 60:880–882. [PubMed: 19564216]
- Grella CE, Greenwell L. Substance abuse treatment for women: Changes in the settings where women received treatment and types of services provided, 1987-1998. *Journal of Behavioral Health Services & Research*. 2004; 31(4):367–383. [PubMed: 15602139]
- Grella CE. From generic to gender-responsive treatment: Changes in social policies, treatment services, and outcomes of women in substance abuse treatment. *Journal of Psychoactive Drugs, SARC suppl*. 2008; 5:327–343.
- Grella CE, Rodriguez L. Motivation for treatment among women offenders in prison-based treatment and longitudinal outcomes among those who participate in community aftercare. *Journal of Psychoactive Drugs. Suppl*. 2011; 7:58–67.
- Guo, S.; Fraser, MW. *Propensity score analysis: Statistical methods and applications*. Sage; Los Angeles: 2010.
- Linacre JM. Rasch power analysis: Size vs. significance: Standardized chi-square fit statistic. *Rasch Measurement Transactions*. 2003; 17:918.
- Linacre, JM. *A User's Guide to WINSTEPS® and MINISTEPS: Rasch-Model Computer Programs: Program Manual 3.70.0*. Winsteps.com; Beaverton, Oregon: 2010.
- Lynch, SM.; DeHart, DD.; Belknap, J.; Green, BL. *Women's pathways to jail: The roles & intersections of serious mental illness & trauma*. Bureau of Justice Assistance, U.S. Department of Justice; 2012.
- Mauer, M.; Potler, C.; Wolf, R. *Gender and justice: Women, drugs, and sentencing policy*. The Sentencing Project; Washington, DC.: 1999.
- Messina N, Burdon W, Hagopian G, Prendergast M. Predictors of prison-based treatment outcomes: A comparison of men and women participants. *The American Journal of Drug and Alcohol Abuse*. 2006; 32:7–28. (2006). [PubMed: 16450640]
- Messina N, Calhoun S, Warda U. Gender-responsive drug court treatment: A randomized controlled trial. *Criminal Justice Behavior*. 2012; 39(12):1539–1558.
- Messina N, Grella CE, Burdon W, Prendergast M. Childhood adverse events and current traumatic distress: A comparison of men and women drug-dependent prisoners. *Criminal Justice and Behavior*. 2007; 34:1385–1401.
- Messina N, Grella CE, Cartier J, Torres S. A randomized experimental study of gender-responsive substance abuse treatment for women in prison. *Journal of Substance Abuse Treatment*. 2010; 38:97–107. [PubMed: 20015605]
- Moos RH, Moos BS. Long-term influence of duration and intensity of treatment on previously untreated individuals with alcohol use disorders. *Addiction*. 2003; 98:325–337. [PubMed: 12603232]
- Mosher C, Phillips D. The dynamics of a prison-based therapeutic community for women offenders: Retention, completion, and outcomes. *The Prison Journal*. 2006; 86(1):6–31.
- McDonald D. Gender-responsive treatment and the need to examine female inmates' lives in prison and prior to prison. *Corrections Compendium*. Nov-Dec;2008 :7–30.
- Ney B, Ramirez R, Van Dieten M. Ten truths that matter when working with justice involved women. *National Resource Center on Justice Involved Women*. 2012:1–18.
- Orwin R, Francisco L, Bernichon T. Effectiveness of women's substance abuse treatment programs: A meta-analysis. *National Evaluation Data Services (NEDS) Analytic Summary*. Jul.2001 21:1–8.

- Orwin RG, Sonnefeld LJ, Cordray DS, Pion GM, Perl HI. Constructing quantitative implementation scales from categorical services data: An example from a multisite evaluation. *Evaluation Review*. 1998; 22(2):245–288. [PubMed: 10183306]
- Pearl, J. On a class of bias-amplifying variables that endanger effect estimates. Technical Report R-356.. In: Grunwald, P.; Spirtes, P., editors. *Proceedings of UAI. AUAI*; Corvallis, OR: 2010. p. 417–424.
- Pelissier BMM, Camp SD, Gaes GG, Saylor WG, Rhodes W. Gender differences in outcomes from prison-based residential treatment. *Journal of Substance Abuse Treatment*. 2003; 24:149–160. [PubMed: 12745032]
- Pelissier B, Jones N. A review of gender differences among substance abusers. *Crime & Delinquency*. 2005; 51:343–372.
- Pelissier B, Motivans M, Rounds-Bryant JL. Substance abuse treatment outcomes: A multi-site study of male and female prison programs. *Journal of Offender Rehabilitation*. 2005; 41(2):57–80.
- Prendergast ML, Messina NP, Hall EA, Warda US. The relative effectiveness of women-only and mixed-gender treatment for substance-abusing women. *Journal of Substance Abuse Treatment*. 2011:336–348. [PubMed: 21315540]
- Robbins CA, Martin SS, Surratt HL. Substance abuse treatment, anticipated maternal roles, and reentry success of drug-involved women prisoners. *Crime & Delinquency*. 2009; 55(3):388–411.
- Rosenbaum P, Rubin DR. The central role of propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
- Rubin DB. Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*. 1997; 127:757–763. [PubMed: 9382394]
- Sacks JY, McKendrick K, Hamilton Z. A randomized clinical trial of a therapeutic community treatment for female inmates: Outcomes at 6 and 12 months after prison release. *Journal of Addictive Diseases*. 2012; 31:258–269. [PubMed: 22873187]
- Salisbury EJ, Van Voorhis P. Gendered pathways: A quantitative investigation of women probationers' paths to incarceration. *Criminal Justice and Behavior*. 2009; 36(6):541–566.
- Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin; Boston: 2002.
- Spjeldnes S, Goodkind S. Gender differences and offender reentry: A review of the literature. *Journal of Offender-Rehabilitation*. 2009; 48:314–335.
- Tang Z, Claus RE, Orwin RG, Kissin WB, Arieira C. Measurement of gender-sensitive treatment for women in mixed-gender substance abuse treatment programs. *Drug and Alcohol Dependence*. 2012; 123(103):160–166. [PubMed: 22138537]
- Van Voorhis P, Wright EM, Salisbury E, Bauman A. Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*. 2010; 37(3):261–288.
- Van Voorhis P. On behalf of women offenders: Women's place in the science of evidence-based practice. *American Society of Criminology*. 2012; 11(2):111–145.
- Wells D, Bright L. Drug treatment and reentry for incarcerated women. *Corrections Today*. Dec.2005 : 98–111.
- Wright BD, Linacre JM, Gustafson JE. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994; 8:370.

Table 1

Maximum likelihood estimates of hazard ratios of fixed effects of predictors on arrest outcomes

Predictors	Overall Arrests	Drug-Related Arrests
GS treatment	1.02	0.89 [†]
Propensity score	0.97	1.00
Pre-treatment arrests	2.59 *	3.02 *
Men's hazard ratio	3.20 *	3.46 *
Controlled environment	0.53 *	0.37 *
Treatment completion	0.89	0.71 *
GS by treatment completion	0.89 *	--

Notes.

[†]
 $p < .10$ *
 $p < .05$.

Table 2

Odds ratios of fixed effects in 3-level outcome models

Fixed Effect	Overall Arrests	Drug-related Arrests
Pre-treatment period intercept	0.08 *	0.02 *
Propensity score	1.06	1.03
Pre-treatment period slope	1.21 *	1.29 *
Propensity score	0.99	0.98
Post-treatment period intercept	0.54 *	0.73 †
Propensity score	1.03	0.94
GS treatment	1.08	0.84 *
Treatment completion	0.81	0.71 *
GS treatment by treatment completion	0.84 *	--
Post-treatment slope	0.91 *	0.83 *
Propensity score	0.98 †	1.01
Controlled environment (time varying)	0.98 *	0.97 *
Men's arrest rate (time varying)	1.00	0.97 *

Note.

†
 $p < .10$ *
 $p < .05$.