CHAPTER 8: ANALYSIS AND RESULTS

This chapter describes the approaches taken to model the four types of post-release outcomes for the subjects in our drug treatment evaluation study and presents the results of these models. The four types of outcome measures are as follow: recidivism, drug use, post-release employment, and failure in a Community Corrections Center (CCC). The chapter begins with a description of the outcome measures. The next section provides a description of our three different analytic strategies. This is followed by a section describing the procedures we used to handle missing data.

After we describe missing data procedures, we discuss results for the two primary outcome measures: recidivism and drug use. Following these results, we present results for two other types of outcomes, employment and failure in a CCC. For each type of outcome measure, we first describe the diagnostic tests, when applicable, and the steps taken in response to these tests. We proceed to compare the treatment effect across the various analytic strategies. Lastly, we highlight gender differences in the other predictor variables found to be significant. Appendix A contains a comprehensive codebook of the variables used in the different analyses.

Outcome Measures

We report on four outcomes: recidivism, drug use, employment, and failure in a CCC. With the exception of our outcome measure for failure in a CCC, the primary source of information for post-release outcomes was telephone interviews with Probation officers. The automated BOP SENTRY database was used to obtain the outcome information on CCC failures.

Probation officers only supervise offenders who have either been sentenced to supervised released or who have been released from prison before completing their entire terms. Approximately 86 percent of the research subjects – 87 percent of the men and 84 percent of the women – were released to supervision. Thus, for our outcome measures of drug use and employment, the analyses were limited to the subjects who were released to supervision. We were able to obtain arrest information for subjects not released to supervision from the NCIC database.[1] Thus, although the focus of our outcome analyses is upon supervised subjects, we were also able to report arrest for new offense results for all subjects.

Data collection from Probation officers was scheduled for three points in time following a subject's release: at 6 months, 18 months, and 3 years. If an individual terminated supervision

[1] Post-release data from a Probation officer were not obtained for 21 individuals released to supervision because we could not identify and locate the supervising officer or did not obtain a response from the Probation officer. For these subjects, arrest data were obtained from the NCIC database.

within 3 years of release, the information from Probation officers was collected through the end of supervision.

Recidivism

Our measure of recidivism was defined as the time from release until there was an arrest for a new offense or supervision was revoked. We included revocation for a technical violation of the conditions of supervision in our definition of this measure, although technical violations have nothing to do with an arrest for a new offense because it was a *competing event*. Unless the competing event was independent of an arrest event, the parameters associated with survival analysis – the type of analysis used for most of our outcome measures – would have been biased and inconsistent. Because similar underlying processes, such as a return to drug use, can trigger an arrest for both a new offense and a technical violation, assuming independence may have been unwarranted. One way to deal with this problem was to treat the criterion variable as either an arrest or a revocation, and that is what we chose to use as our primary approach to measuring recidivism.

Despite our focus on measuring recidivism as an arrest or a revocation, we also provide results limiting the outcome measure to arrest for a new offense. These additional results are provided because our primary interest was in arrest for a new offense, because it indicates more of a risk to public safety than a typical supervision revocation. Furthermore, we also report results based on a new offense for all subjects, both supervised and unsupervised. We did this to ascertain whether the supervision process itself either affected behavior or what was observed about behavior.

Arrest data obtained from Probation officers contained all arrests during supervision, regardless of whether the individual was convicted or incarcerated. The measure of arrest for this report was defined as the first occurrence of an arrest within 3 years after release from custody. We analyzed the single most serious offense for individuals with multiple charges at the time of the first arrest.[2]

Some individuals had state detainers immediately upon release from BOP custody or at some other point during their supervision by a Probation officer. For these individuals, time spent on a detainer was excluded from time supervised. Thus, an individual who spent time in prison for a detainer immediately upon release from BOP custody was assigned a release date which corresponded to the release from the detainer.

We verified the consistency of information on arrests for a new offense by comparing Probation

---

[2] NCIC codes are ordered by severity of offense, so we used their hierarchy when determining which offense was to be considered the most severe.

officer arrest information to NCIC arrest information. This comparison was done for 50 randomly selected subjects with arrest information obtained from a Probation officer. We found only one subject for whom NCIC data showed an arrest not reported by a Probation officer,[3] which gave us reason to believe in the comparability of data from the two sources (*i.e.*, Probation officers and NCIC).

Although we attempted to obtain 3-year follow-up data, such data were unavailable for some subjects. There were 141 subjects for whom there were only 18-month follow-up interview data from Probation officers: 91 treatment subjects and 50 comparison subjects (44 women and 97 men). This was due to the lack of available staff to complete these interviews. In addition, 50 subjects died during the 3-year follow-up period. Lastly, some individuals completed supervision within less than 3 years: 12 subjects completed supervision within 12 months and 30 subjects completed supervision within 24 months. For administrative reasons, we were also unable to obtain NCIC data on 43 of the 374 offenders that were not released to supervision.

Drug Use

Drug use information was obtained from the Probation officers' records of violations of conditions of supervision for our research subjects and includes illicit drug use as well as alcohol use. Four different violation categories were used as indicators of drug or alcohol use: a positive urinalysis test for any illegal drug, a refusal to submit to a urine test, a positive breathalyser test for alcohol, and an admission of drug use to the Probation officer. When a person refused a urine test, the assumption was that he or she would have had a positive urine test result. We limit this outcome measure to individuals who 1) were released to supervision and 2) were tested for drug use. There were 150 men and 43 women released to supervision who were not included in the analyses of drug use because they had no urinalysis testing.[4] Although we would have ideally liked to model the *number* of drug use occurrences, we limit our analyses to the first occurrence. This was necessary because in some districts an individual is revoked after the first positive urinalysis while in other districts individuals are revoked only after repeated positive test results.

---

[3] The Probation officer was recontacted to verify that we had correctly recorded the arrest information.

[4] Two of the individuals who did not receive urinanalysis testing during supervision were detected as having used alcohol (by means of a positive breathalyzer or admission of alcohol use to the Probation officer). However, information on alcohol-only users is insufficient for inclusion in drug use models because the majority of detected drug use is for an *illicit* drug (*i.e.*, there is no regular screening for use of alcohol). Future analyses will assess why some individuals did not receive urinalysis testing.

Employment

Although the primary focus of the Bureau's drug treatment programs is upon illegal behavior and drug use, employment is also considered an important outcome measure. Lack of employment has been shown to be related to criminal activity (Horney, Osgood, and Marshall, 1995; Crutchfield and Pitchford, 1997; Fagan and Freeman, 1999; Uggen, 1999; Cohen et al., 2000). Furthermore, employment is an indicator of the prosocial behavior which the drug treatment program seeks to encourage. Probation officers provided information about the starting and ending dates — as well as the number of hours worked per week — for each job the subject held during supervision. With this information, we created two different measures of post-release employment. The first consisted of the percentage of time employed during the follow-up period. For each week of post-release supervision a value of 40 hours of available work time was allotted. The percentage reflects the actual number of hours worked during the supervision period divided by the number of hours available. For individuals who worked more than 40 hours per week during the entire time period, values were truncated to 40 hours. Thus, the values for this measure of employment ranged from 0 to 100 percent. The time detainees spent incarcerated after release from BOP custody for a previous offense (*i.e.*, an offense committed before their most recent admission to the BOP custody) was excluded as time available for employment.

A second criterion was a categorical variable representing the level of employment during the follow-up period. This second measure was applied only to people who were members of the workforce. We excluded those who were retired, disabled, in school full-time or had full-time child care responsibilities. The categories were as follow: employed full-time during the entire post-release period, employed full-time during some portion of the post-release period, employed part-time during some or all of the post-release period, and not employed during the post-release period. As with our previous employment measure, time spent on a detainer for a previous offense was excluded as time available to work.

Failure in a Community Corrections Center (CCC)

Our fourth outcome measure was intended to detect whether or not treatment improved performance during a Community Corrections Center stay. Many BOP offenders go through a transition period between prison and community supervision during which they are placed in a halfway house. While in the halfway house, they are at liberty to work or go to school during the day, but they must return to the house when not occupied with approved activities. Some offenders spend the latter part of their CCC placement on home confinement with electronic monitoring. Not all BOP offenders serve time in a CCC however, so this analysis was restricted to those who received a CCC placement.

Analytic Strategies

As discussed in our previous 6-month preliminary report, selection bias was the most significant methodological problem faced in studies evaluating the effect of drug treatment programs. This section provides an overview of the methodological problem and the approach we took to address this problem (*see* Chapter 2 for a more detailed discussion of this methodological problem).

The randomized experimental design is the "gold standard" for evaluation research. The simplest version of this design requires that members of an eligible population be randomly assigned to either a treatment group or to a control group. Provided that other factors do not contaminate the experiment, comparing the outcomes for the treated group and the untreated group provides an unbiased measure of the average treatment effect.

Despite its appeal, a randomized design may be impractical in some settings, such as criminal justice populations, where due process or other administrative constraints restrict randomization of otherwise equivalent populations to treated and untreated conditions. Even when implemented, randomized experiments often collapse as agencies thwart researchers' evaluation plans or subjects refuse to cooperate. Much of what researchers know (or think they know) about treatment programs comes from evaluations based on quasi-experimental designs. A quasi-experimental design typically uses statistical controls in place of random assignment to establish an assumed equivalency between a treated group and a (generally) nonequivalent comparison group. The statistical control is sometimes compelling, but rarely convincing, because it does not transform *association* (treated subjects tend to have better outcomes) into *causation* (treatment causes better outcomes), as randomization does. Quasi-experimental designs invariably end with the caveat: "These findings might represent a treatment effect, but we cannot be sure because …" Still, not all quasi-experiments are created equal. Some have a long list of caveats, while for other quasi-experiments, the qualifications might be relatively innocuous. Indeed, a well-designed quasi-experiment can provide strong evidence for rejecting a null hypothesis that a program has no appreciable treatment effect.

This evaluation was a quasi-experiment because we could not randomly assign inmates who abused substances to treated and untreated conditions (*see* Chapter 5). As is true of most substance abuse treatment outcome evaluations, the principal analytic problem was how to deal with potential selection bias. As discussed in Chapter 2 selection bias has been recognized in program evaluation literature and sociological literature but is not widely recognized within the drug treatment evaluation field.

Economists and others have used selection bias adjustments for a long time, but the flurry of research applying this approach to quasi-experiments is more recent. In the late 1970s, Heckman

(1979) developed an influential approach for dealing with selection bias that some researchers took to be a solution, at least within the context where it could be used. Here, we refer to that approach as "Heckman-type" adjustments. Unfortunately, subsequent research has shown that Heckman's solution rests on strong distributional assumptions, and results are sensitive to getting those assumptions right (for example, LaLonde, 1986). This would be no problem if the assumptions were testable, but in many cases they are not, or else the test lacks power. In his influential paper, LaLonde (1986) demonstrated that any quasi-experiment using Heckman's approach to control for selection bias could yield estimates of the treatment effect that suffered from large biases. Some methodologists may even have regarded LaLonde's demonstration as the end of quasi-experimental design as a method for evaluating treatment programs (*see* Burtless, 1995). Such an assessment would be premature, because methods for dealing with selection bias continue to evolve (Manski and Nagin, 1998).

Heckman and his colleagues (for example, Heckman and Smith, 1995) have argued that LaLonde overstated the case against dealing with selection bias encountered by quasi-experimental design. Whatever the merit of their case, LaLonde's paper galvanized the development of alternative ways of dealing with selection bias. Recent theoretical expositions include Smith (1997), Heckman et al. (1998), and Dehejia and Wahba (1999).[5] Those recent papers have stimulated our own approach to dealing with selection bias in quasi-experimental design. While there is no agreement about the best way to address selection bias issues, it has been recommended that researchers use a variety of methods (Winship and Mare, 1992). Such an approach can be viewed as a form of sensitivity analysis and provides some evidence a researcher can use to assess the extent to which findings are method-dependent.

As mentioned previously, the BOP was unable to assign subjects randomly to treatment and no treatment conditions. Thus, a quasi-experimental design was used to test for treatment effectiveness. Some prisons had residential drug treatment programs (DAP) and others did not. Inmates in DAP facilities did not differ materially from prisoners in non-DAP facilities, so the two populations were comparable for evaluation purposes. Within the DAP facilities, some offenders were offered and accepted treatment (hereafter the **DAP *treatment group***) while others either were not offered treatment or declined treatment that was offered (hereafter the **DAP *comparison group***). Of course, those offenders who were housed in non-DAP facilities did not receive treatment (hereafter the **non-DAP *control group***).[6]

---

[5] At least one approach, the use of instrumental variables, predates La Londe (*see* the discussion in Maddala, 1983, Chapter 9). Nevertheless, our impression is that attention to this approach has accelerated since LaLonde's paper.

[6] *See* Chapter 5 for discussion of how we selected and identified DAP comparison and non-DAP control subjects.

As discussed in our 6-month preliminary report, a simple comparison of the outcomes for offenders who were treated (the DAP treatment group) with the outcomes for offenders who were not treated (the non-DAP control group and the DAP comparison group) could be misleading because of selection bias. In this case, the concern was that some unmeasured factors (such as motivation to change) that affect the decision to enter treatment might also affect post-release performance, so the relationship between treatment and post-release performance could be partly or wholly spurious. In addition to including control variables in a regression model, which we will refer to as the *unadjusted* approach, we adopted two other analytic methods for dealing with selection bias: a *standard instrumental variable approach* and a *Heckman selection bias approach*.

Unadjusted Approach

Our first approach – the unadjusted approach – uses a dummy variable to represent the treatment effect. This dummy variable identifies whether the individual received and completed treatment or did not. The category of those who did not included individuals who received treatment but did not complete it, individuals who had treatment available but did not volunteer, and individuals who did not have treatment available. Despite some similarity of our unadjusted approach to analytic conventions used in previous studies, we tried to improve upon these previous studies in several ways. First, we included a comprehensive set of control variables. Second, we used event history techniques that most adequately controlled for the right censoring of data (Allison, 1984; Blossfeld, Hamerle and Mayer, 1989).[7] Third, our study was multi-site and this increased the generalizability of our findings.

Instrumental Variable Approach

As discussed in Chapter 2, our first approach to addressing selection bias was the instrumental variable approach. This is the most straightforward of the approaches used. The key is to develop a suitable instrument (Davidson and MacKinnon, 1993, for example). Suppose an analyst was to combine data from all three sources (non-DAP controls, DAP comparisons, and DAP treatment), assign a dummy variable coded one to those who received treatment and coded zero for those who did not, and then regress the outcome variable on this dummy variable and any control variables that seem appropriate. The problem with this approach is well known. The estimated regression parameter associated with the dummy variable will be biased and inconsistent if the dummy variable and the error term are not independent. Independence seems unlikely if any

[7]Right censoring occurred for study participants who had not experienced the post-release outcome in question. These subjects remained at risk of experiencing the event, and our event history models took the censoring of the observations at the 3-year point into account.

unmeasured factor (such as motivation) affects both the receipt of treatment and the outcome variable.

A solution is to identify an instrumental variable that is highly correlated with the treatment variable but that is distributed as independent of the error term. One suitable instrument is the estimated probability of entering and completing treatment, where this instrument might be estimated from a probit model. It should be noted, that the resulting instrument – the probability of completing treatment – the dependent variable in the probit model, is a variable indicating whether or not the offender entered and completed treatment (*see* discussion of equation **(7)** in Chapter 7). Our instrumental variable is independent of the error term because one of the predictors is the probability-of-volunteering coefficient and the amount of time housed at a DAP site at a time when treatment was available. Inmate assignment to institutions during the study period was independent of drug use history. Furthermore, institutions started their drug treatment programs at a different times and this affected the rate of volunteering for treatment. This probit model is estimated using just those data from the DAP subjects, since the non-DAP subjects have a zero probability by definition, so the instrument is set to zero for them.[8] By substituting the instrument (the estimated probability of entering and completing treatment, which we refer to as the probability of being treated) for the dummy variable, and estimating the regression, the parameter estimate associated with the instrument provides an estimate of the average treatment effect that is free of selection bias. The instrument – the estimated probability of entering and completing treatment – used in our outcome analyses was weighted to represent the proportion of the sample that was treated. Our sample selection process was disproportionate: we oversampled treatment participants but undersampled from the non-volunteer (e.g., DAP comparison) subject group.

The instrumental variable approach to evaluating treatment effectiveness is not much complicated by introducing control variables and using regression models. The introduction of control variables has three benefits: By reducing unexplained variance, the regression can reduce the standard error of the estimate for the treatment effect. Second, the control variables can help adjust for any population difference between DAP and non-DAP facilities. And, third, the parameters associated with control variables have policy relevance for the BOP.

Heckman Approach

A second approach to selection bias, called herein the *Heckman selection bias* approach (Heckman, 1979; Maddala, 1983) is somewhat more difficult to apply than is the instrumental variable approach. It requires the analyst to *jointly* model the selection into the sample and the

[8] We discuss the probit model in Chapter 7.

post-release outcome. Here, we note that the selection bias approach has much in common with the standard instrumental variable approach, and if the analyst is willing to limit his analysis to a linear-additive regression model, there is little to recommend the selection bias approach over the instrumental variable approach. However, as explained by Maddala (1983, p.261), the Heckman selection bias models can be used to study more complicated models where treatment interacts with other variables.

Heckman suggested a two-equation estimator. The first equation described the selection process, and the second described the post-treatment outcome. Parameter estimation required identification conditions, typically, that some of the variables that entered the first equation did not enter the second equation. This condition is difficult to satisfy in many practical settings. However, we were able to do so because several of our predictors of treatment entry and retention were not in our second equation. Our outcome equation does not contain information on history of violence, severity level of offense, the Change Assessment Scale, the Hope Scale, and the planful problem-solving subscale of the Ways of Coping survey. In addition, our selection equation contains two predictors -- probability-of-volunteering coefficient and time housed at a DAP site – which are not in the outcome equation and which are not associated with outcome. During the time of our study, individuals were not assigned to treatment sites based upon treatment need.

Modeling Techniques

Survival models were used to perform the analyses of arrests and revocations and of drug use. The planned duration in a Community Corrections Center varied across offenders, so we also adopted a survival model to study the time until failure of CCC placement where successful completion was the censoring event. In survival models we model the time to occurrence of the event of interest (e.g., arrest, drug use, etc.) during the risk period, that is, the time after an individual is released from custody and is under the supervision of a Probation officer. Survival modeling is the most suitable type of analysis because it models not only whether the event occurred, but when the event occurred. In addition, survival analysis is able to handle "censored" data, that is, data for individuals who did not fail during the 3-year post-release observation period, as well as data for individuals who were not observed during the entire post-release period due to termination of supervision before the end of the 3-year period, incarceration for a detainer, or death. We discuss the specific parametric assumptions of the survival models in the "Diagnostics" section of this chapter.

Parametric survival models typically assume that all subjects must eventually recidivate (fail) if given enough time. An alternative assumption is that a proportion (PRO) will recidivate given a follow up period of infinite length, but 1-PRO will never recidivate. This is referred to as a split-

population model. We felt that it would be unreasonable to assume that all subjects would eventually fail. Therefore, we modified the likelihood function to accommodate a split-population assumption (*see* Greene, 1998).

We assumed that:

$$PRO = \frac{1}{1 + e^{-S}} \qquad \textbf{(8)}$$

where S is a parameter to be estimated and defines whether the subjects is or is not in the failure group. The parameter S is reported in all our survival models.

As mentioned in our section on outcome measures, we had some subjects with 18-month and not 3-year follow-up data. In addition, some subjects died during supervision and others completed supervision in less than 3 years. Event history techniques allowed us to take these different risk periods into account by censoring data when we were not able to obtain 3-year follow-up data for all subjects.

We used a two-limit tobit model (0 percent employment and 100 percent employment limits) to test the treatment effect for our first measure of post-release employment — employment rate. Because our second indicator of employment consisted of level of employment which was coded on an ordinal scale, we used an ordered probit model.

Appendix B provides a technical explanation of the Heckman-type statistical models used in this study. There are four versions of the Heckman model:

1. a lognormal survival model
2. an exponential survival model
3. a probit model
4. a two-limit tobit model

Each of these models has an adjustment for selection bias.

Presentation of Results

For the recidivism and drug measures, we used three approaches. These approaches, as discussed earlier in this chapter, are the unadjusted, the instrumental variable and the Heckman-type. For one of our measures of employment – employment level – and for halfway house failure, we used only the unadjusted and the instrumental variable approaches.

We modeled male and female outcomes separately. As discussed in our review of the literature on gender differences (*see* Chapter 3), the process of change from a drug using criminal lifestyle to one without drug use and criminal activity may differ between men and women. Furthermore, men and women were in separate treatment programs. A thorough representation of male and female differences would have required the inclusion of a large number of interaction terms in analyses of men and women combined.

In presenting the results, we discussed coefficients significant at $p <= .05$. We used a two-tailed test for all coefficients except those representing effects for DAP in-prison treatment. For these treatment effects, we used a one-tailed test because we hypothesized that individuals who received in-prison drug treatment would have more favorable outcomes than would those who did not receive drug treatment. Furthermore, as argued by Lipsey (1990), evaluation research is better served by accepting an increased likelihood of a Type I error and lowering the probability of a Type II error. Unlike basic research, where it is more desirable to take a conservative approach to making conclusions about relationships between various phenomena, evaluation research may be better served by decreasing the probability that an effective treatment is falsely found to be ineffective (Type II error).

Although there is considerable overlap in predictors of our outcome measures, we chose predictors specific to each outcome measure. For example, never having a period of 30 or more days of unemployment before incarceration was used as a predictor only in our models of employment. History of drug and alcohol treatment were used only in models of drug use.

Our predictor variables can be separated into those that have a substantive interpretation and others that serve simply as control factors for which we do not provide any interpretation. These control factors represent different levels of supervision or services that individuals received which are determined by prison policies, self-selection, or Probation officers. The control variables used in one or more of our outcome analyses included: participation in vocational education during incarceration, percent of time employed in prison industries (UNICOR) during incarceration, Community Corrections Center placement, frequency of urinalysis testing during post-release supervision, frequency of personal and collateral contacts with Probation officers during supervision, participation in a self-help group after release, assignment to receive treatment while under post-release supervision, and not being in the work force after release from BOP custody. Because the control variables do not have a substantive interpretation, we chose not to discuss those coefficients.

Effects vector coding was used for most nominal and ordinal level variables (*see* codebook in Appendix A). In effects vector coding, each coefficient represents the contrast of that category with the adjusted grand mean. Dummy variable coding was used only in one case when we had a referent category of particular interest. We used dummy variable coding for the categories of

daily drug use in the year before arrest, since we were interested in the contrast between the various types of drugs used on a daily basis when compared to no daily drug use in the year before arrest.

We focus the discussion of the results, as well as the conclusions that follow, on the effects of the in-prison DAP. Our comparisons are directed towards identifying the consistency of the effects of DAP treatment across the different models. In discussing the other predictor coefficients that are significant, we chose to discuss the significant coefficients in the Heckman model. This is the approach which we feel represents our "best" model because it identifies whether selection bias occurred, and if so, the direction of the bias. For our CCC failures, where we did not use the Heckman model, we report the significant coefficients in the instrumental variable approach. We note that with the exception of rare instances, those predictors which are significant using one analytic technique are significant in all other models having the same outcome measure.

Our discussion of the other predictor variables focuses upon gender differences. As mentioned earlier, we modeled men and women separately in order to learn whether and how women may differ from men in their post-release outcomes.

Missing Data

We eliminated 18 subjects from our analyses because these individuals did not have information on the probability of volunteering for and completing treatment.

We decided to handle missing values for our predictor variables by imputing values rather than accepting the default method of listwise deletion for the analyses. Without using these methods, listwise deletion would have resulted in too many cases being omitted from the analyses. We felt that the potential biases resulting from a listwise deletion of a large number of cases would be greater than those resulting from imputation. In addition, deletion of subjects with missing data has been shown to result in less efficiency (Little, 1982; Little and Rubin, 1987; Rubin, 1987; Schafer, 1997; King, 1999). In our preliminary report (Pelissier et al., 1998), we tested whether the use of imputed values modified the results. We found no significant differences in the coefficients representing the effects of treatment when comparing results using listwise deletion to those using imputed values for missing data.

We estimated categorical and ordinal values with a hotdeck procedure and continuous values with BMDP's (Dixon, 1992) maximum-likelihood missing values procedure. The following were the categorical and ordinal variables and the number of cases for which values were estimated:

- Employment status in month before incarceration (37);

- Previously supported self at least one year mainly through illegal sources (n=15);
- Unemployed 30 days or more at least once in lifetime (n=104);
- Daily drug use during year before arrest (109);
- Spouse ever had a drug problem (n=26);
- Psychiatric diagnoses of antisocial personality or depression (n=213);
- Previous mental health treatment (n=6);
- With whom living after release (n=251);
- Treatment received after release (n=20);
- Self-help group participation after release (n=20).

Five continuous variables were estimated using the maximum-likelihood regression procedure in BMDP. These variables were:

- Age at first commitment (n=8);
- Level of education (n=5);
- Average number of contacts with Probation officer during the first 6 months of supervision (n=7);
- Average number of monthly collateral contacts of Probation officer during the first 6 months of supervision (n=7);
- Average number of urinalysis tests during the first 6 months of supervision (n=14).[9]

Model Specification Diagnostics

We performed several diagnostics for testing model specification for all outcomes except employment, the only model where we did not model time to failure. The first test was to fit three alternative versions of parametric survival models based on the lognormal, exponential, and Weibull distributions. We selected the distribution with the highest likelihood as the "best model" because it provided the best fit to the data.[10]

Of course, the model with the highest likelihood is not necessarily a good model. The second diagnostic test was to inspect plots of the integrated hazard for the selected model on the horizontal axis against minus the logarithm of the integrated hazard on the vertical axis. This approach was recommended by Lancaster (1990, page 312). If the model is a satisfactory one,

---

[9]Although substitution for missing values tends to bias parameter estimates toward zero, we did not consider this problematic as such substitution did not occur frequently.

[10] We used the instrumental variable approach in the first two diagnostic tests of the survival models

then the plot falls on a 45-degree line. We judged whether or not the model was acceptable by inspection. The plots are presented in Appendix C.

A third test was to compare the parameter estimates for the treatment effect provided by the two models we used to correct or minimize selection bias: the instrumental variable model and the Heckman-type adjustment model. Both should yield similar, but not necessarily identical estimates. If the estimates are not similar, then one must be suspicious of the distribution assumptions made about the mixture distribution adopted in the Heckman-type adjustment model.

The first indicator of recidivism was arrest for a new offense for all subjects. This criterion measure included 2,099 men and 547 women in the analysis. Using the instrumental variable approach, we estimated survival models based on the lognormal, the exponential, and the Weibull distributions. Table 32 reports the minus log-likelihood values. The value closest to zero denotes the best model. On the basis of that test, we selected the exponential as the best survival distribution for men and the log-normal as the best survival distribution for women.[11]

The second test plots the integrated hazard against minus the logarithm of the integrated hazard. Plots, based on the best model as determined by the likelihood comparison, appear in Figures 1 and 2 (*see* Appendix C).

For men, the diagnostic test leads to the conclusion that the exponential is a suitable distribution for modeling the time until an arrest. For women, however, the test calls the distributional assumption into question. The model based on the log-normal performs well for the early part of the integrated hazard, but it does not perform well for the latter part. This observation will have consequences for the analysis.

We tried two strategies to address the apparent problem in using the log-normal distribution to represent time to failure for women. First, we censored the follow-up period at 12 months and at 18 months to see if any of the three distributions worked better over a shorter span. They did not; the same diagnostic problem persisted. The second approach involved combining the instrumental variable approach with a Cox proportional hazard model. The Cox proportional hazard model makes no distributional assumptions. We discuss the results of this strategy when we discuss the treatment effects.

---

[11]The model based on the Weibull will always have a larger likelihood than the model based on the exponential, which is a special case of the Weibull. Unless the Weibull was significantly better than the exponential, we adopted the exponential.

Table 32 once again shows the results from the first diagnostic test to identify the distribution which provides the best fit to the data. Those tests caused us to again select the exponential as the best way to represent the failure time for men and the log-normal as the best way to represent the failure time for women. The second diagnostic, the plots based on the integrated hazard, was similar to the plots for men shown above, so we do not show new plots here. For women, the plot of the integrated hazard (*see* Figure 3, Appendix C) is much improved, suggesting that the lognormal is an acceptable failure time distribution. As mentioned earlier, using an arrest as a criterion of failure is problematic because revocation for a supervision violation is a competing event. That would not change the way we look at the problem if an arrest and revocation could be treated as stochastically independent, but an assumption of independence seems unjustified. Our next model treats the outcome as either an arrest or a revocation.

Diagnostic tests for this new model again lead us to adopt a survival model based on the exponential distribution for men and a model based on the lognormal distribution for women (*see* Table 32). The plots of the integrated hazard were similar to the previous plots, so we do not show them.

Next we analyzed the time until relapse to drug use. We could only do this for study subjects who were supervised and had their urine tested as a condition of supervision. There were 1,692 males and 430 females. Diagnostic tests (*see* Table 32) suggested that the lognormal model was better than the exponential model for both men and women. For women, the Weibull model was slightly better than the lognormal. The difference was slight, however, and given that we had not developed a Heckman-type adjustment correction for the Weibull, we adopted the lognormal.

The second diagnostic, based on the integrated hazard plots, seemed to show that the assumption of the log-normal distribution was acceptable for men (*see* Figure 4, Appendix C). However, we encountered the familiar problem for women. The model "worked" in the range of lower integrated hazard scores but not when the scores got larger (*see* Figure 5, Appendix C).

The final outcome considered in this study for which we used survival analyses is failure during a Community Corrections Center stay. A total of 1,476 men and 409 women entered halfway houses, and those men and women were the subject of this analysis. The outcome variable was failure for any reason during that placement.

Diagnostics indicated that neither the log-normal nor the exponential were suitable candidate distributions for the survival analysis, because each was clearly dominated by the Weibull distribution (*see* Table 32). The dominance of the Weibull distribution caused some problems, because we had not written a version of the Heckman-type adjustments based on the Weibull distribution. Instead, we used the routine programmed in LIMDEP (Greene, 1998), but we only estimated the unadjusted and the instrumental variable model.

The parameter estimates for the treatment effect for all of the different dependent variables and the t-scores for those estimates are summarized in Table 33. The treatment effects are represented separately for men and women. Tables 34 to 47 show the coefficients for all of the other variables in each of the analyses.

Results

Recidivism

We present our results for recidivism beginning with our first criterion measure — arrest for new offense for all subjects. We continue with a discussion of our second criterion — arrest for a new offense for supervised subjects — and conclude with the criterion we consider to be the most desirable criterion — arrest for a new offense or supervision revocation for supervised subjects. Tables 34, 36, and 38 present the results for men for each of the three criterion measures of recidivism and Tables 35, 37, and 39 the results for women. These tables contain the parameter estimates and t-tests for all the variables in the model including the variable testing the treatment effect; however, readers can also refer to Table 33 for a summary of the treatment effect for each measure of recidivism and for each of the different analytic approaches. Because the sign of the coefficient indicative of a positive treatment effect varies with the distribution of the survival function, this Table also identifies the underlying distribution for each survival model and the sign – positive or negative – which is associated with a positive treatment outcome.

In tables 34 to 39, WASTREAT represents the treatment effect. Tables 34 to 39 contain the parameter estimates and t-scores for every variable used in each of the analytic approaches -- unadjusted, instrumental, Heckman. For the Heckman model, the variable ν represents the estimate of the selection bias coefficient. Definitions for all other predictor and control variables are found in the codebook in Appendix A. Since our most desirable criterion is arrest or revocation, we limit our discussion of the effects of predictor variables other than the treatment effect to this criterion. We note, however, that there is one predictor – supervision versus no supervision after release – which is included in the model where the criterion variable is arrest for a new offense and where we included all subjects in the model. This is the only model where we include both supervised and unsupervised subjects. Therefore, we do discuss the effects of this variable in the recidivism model for all subjects.

<u>Arrest for a New Offense - All Subjects</u>

Thirty-five percent of male subjects, including both those supervised and those unsupervised, were arrested for a new offense. In looking at the coefficient WASTREAT in Table 34 we see that all three modeling strategies indicate that for men the treatment effect was statistically significant. Unlike the log-normal or Weibull distribution where a positive sign indicates

increased survival time, a negative parameter denotes a longer survival time (reduced likelihood of an arrest) in the exponential model. Two other findings were important, however. The first was that the instrumental variable approach and the Heckman-type adjustment approach produced parameter estimates that were larger than the estimate for the unadjusted model. The second was that the two methods used to adjust for selection bias yielded estimates that were roughly consistent with each other, although they were not identical. The treatment parameter was larger in the Heckman model than in the instrumental model. Clearly an analyst should not be indifferent toward controlling for selection bias in this context.

Approximately twenty percent of all the women were arrested for a new offense. Table 35 shows that for women, none of the three approaches suggest that treatment was effective at reducing criminal recidivism. The parameter estimates in Table 35 had the expected signs (unlike the exponential model, positive denotes a favorable treatment effect in the log-normal model), but none approached statistical significance. Perhaps treatment did not work for women in reducing recidivism, but we have to be suspicious of the fact that, while the log-normal was the best of the three distributional assumptions maintained in this study, Figure 2 (*see* Appendix C) showed that the log-normal was not especially descriptive of recidivism for women. As mentioned in our discussion of diagnostic tests, the same diagnostic problems persisted when we censored the follow-up period at 12 months and 18 months. Thus, we tried another approach. We combined the instrumental variable approach with a Cox proportional hazard model, which does not impose any distributional assumptions. It does, of course, impose restrictions on the hazards. The resulting t-score was only –.06. Consequently, we conclude that treatment effectiveness has not been demonstrated for women, at least when arrest for a new offense was the criterion and the entire release sample was used.

Arrest for a New Offense - Supervised Subjects Only

As mentioned earlier, using arrest as a criterion is problematic when analysis is based on all offenders because some were not released to supervision. Analyses for our second indicator of recidivism, arrest for a new offense for supervised subjects, included 1,842 men and 473 women. Approximately 33 percent of the men and 17 percent of the women were arrested. Table 36 shows that for men, the treatment effect would be judged as statistically significant using our one-tailed test. The models that account for selection bias increase the size of the treatment effect parameter but that finding would seem to be inconsequential given the small values for the t-scores. For women, the results in Table 37 agree that there is no significant treatment effect.

Arrest for a New Offense or Revocation - Supervised Subjects Only

For our third criterion measure — arrest for a new offense or revocation among supervised subjects — approximately 55 percent of the men failed as compared with 34 percent of the women. Analyses for arrest or revocation for supervised subjects shows that all three methods of estimating the treatment parameter agree that the treatment effect was statistically significant for

men (*see* Table 38). The two methods used to adjust for selection bias yielded roughly similar parameters, which were larger than the treatment effect estimated in the unadjusted model. In fact, once we controlled for selection bias, the treatment effect was more than double or triple what we would have otherwise estimated.

For women, the three approaches agreed that substance abuse treatment did not seem to improve the post-release outcomes, at least when those outcomes were judged by an arrest or revocation (*see* Table 39). When we examined the results by censoring the follow-up period at 18 months, the parameter estimate also was not statistically significant (t=0.33). A Cox proportional hazard model lead to the same findings. This lead us to infer that the treatment effect for women was not large and that model misspecification was probably not the explanation.

It is worthwhile to note how the parameter estimates associated with the treatment effect varied by estimation method. The three approaches yielded different estimates for men, and the treatment effect was stronger when estimated by methods that controlled for selection bias. The Heckman-type adjustment model helped explain the reasons for these differences. As we already noted, the variable ν is an estimate of the selection bias effect. In looking at the coefficient ν for men in Tables 34, 36 and 38, we see that it is significant only in Table 38 — arrest or revocation for supervised subjects. The positive sign of the coefficient indicates that men who entered and completed treatment were more likely to fail, when the criterion measure was arrest or revocation. Thus, by accounting and adjusting for the selection of riskier men into treatment programs, the treatment effects were found to be more influential than they would have been without this adjustment. That explains why the unadjusted model had coefficient estimates indicating less of an effect than the Heckman model. In the unadjusted model, the impact of treatment was muted by the fact that men more likely to recidivate were selected into treatment.

The Heckman-type adjustment model also provided an estimate of the correlation of the error terms that affect the selection into treatment and the failure rates, respectively. Table 40 summarizes estimates of those correlations and reports t-scores for those estimates for all of our outcome measures. The likelihood functions used a transformation, so the correlation is represented as:

$$\rho = 2 - \frac{1}{1+e^{\nu}} \qquad \textbf{(9)}$$

where ρ is the correlation, which is reparameterized in terms of ν, the parameter actually estimated by the method of maximum likelihood. Allowing ν to vary freely, ρ is constrained to

fall between 0 and 1. Table 33 reports point estimates and t-scores for ν and Table 40 reports ρ as derived from the estimated ν.

Our experience with selection bias models suggests that estimates of ρ typically have high standard errors. One should probably not take a lack of statistical significance to mean there was no selection bias; even when ρ lacks statistical significance, the correlation can still affect estimates of the treatment effect.[12] Nevertheless, for men the estimate of ρ was significant in the regression for arrest or revocation and it seems to be sizeable. The direction of the correlation suggests that the worst risks — those most likely to be rearrested — were most likely to enter and complete treatment. Therefore, the models that adjust for selection bias tended to estimate a stronger treatment effect than did models that did not adjust for selection bias. Furthermore, even when ρ was not significant, it still obfuscated the treatment effect..

For women, on the other hand, the estimated correlations for the three criterion measures of recidivism were never large and they never approached significance. This explains why for women the models that adjusted for selection bias gave parameter estimates that were very similar to models that did not adjust for selection bias.

Although Table 33 reports estimates and tests for significance of the parameters associated with the treatment effect, because the statistical models are nonlinear, the parameters are difficult to interpret. We translate those estimates for the treatment effect parameters into metrics that are easier to understand.

We substituted the mean value for each variable that entered every regression. Of course, these means varied from regression to regression, because each regression used a somewhat different variable set and data.[13] Using those means, together with the parameter estimates reported in the tables, we computed the probability of failure for the "average" offender within three years of release from prison. All calculations were based on the instrumental variable model.[14] Those probabilities for each outcome measure are reported in the third column of Table 41 under the heading "Overall."

---

[12] One might treat the t-score associated with the parameter ρ as a test of the null hypothesis of no selection bias. The problem with this approach is that it can lead implicitly to acceptance of the null hypothesis, which has no justification under statistical theory.

[13] The variable set differs primarily because post-release predictors cannot be used in models of arrest for new offense for all subjects where post-release information is not available.

[14] These estimates should be similar for the Heckman model.

We repeated this calculation after substituting a zero in place of the mean treatment effect. This provided an estimate of the probability of failure for someone who was not treated. Then we replicated the calculation after substituting a one in place of the mean treatment effect. This provided an estimate of the probability of failure for someone who was treated. The untreated estimate appears in column four; the treated estimate appears in column five.

As already mentioned, treated males tended to recidivate at lower rates than untreated males. The estimates varied with the model specification. For an arrest, regardless of post-release supervision status, an estimated 38 percent of the untreated group would recidivate compared with 31 percent of the treated group. For an arrest, limited to males who were supervised after release, an estimated 35 percent of the untreated group would recidivate compared with 30 percent of the treated group. For an arrest or revocation, given post release supervision, an estimated 53 percent of the untreated group would recidivate compared with 44 percent of the treated group. Of course, judgment is subjective, but these appear to be sizable treatment effects.

For women, we found no statistically significant treatment effect for arrest. Consistent with that, treated and untreated females were arrested at about the same rates. This is true when we looked at arrests for all subjects as well as arrests for supervised subjects only. We found no significant effect for treatment when an arrest or revocation was used as the criterion variable either. The point estimate, based on the "average" offender, implied that 30 percent of the untreated women would recidivate compared with 25 percent of the treated women. Perhaps a larger sample would have shown a statistically significant treatment effect for women, but we cannot know for sure. We will come back to the drug relapse effects represented in Table 41 when we discuss those models.

We next examine the other covariates which were significant and compare the results for men with those for women. As previously mentioned, we focused our attention on the models where recidivism was defined as an arrest or revocation for those offenders who were subject to post-release supervision. We have argued that this is the most meaningful model since we make no assumptions about whether or how revocations compete with arrests. We limit our discussion of the coefficients in the model of new arrests for all subjects to the coefficient representing supervision status upon release. The variable ESUPRLNO represented in Tables 34 and 35 tested whether supervision had an effect upon post-release arrest. This variable was not significant in either the model for men or women indicating release to supervision had no impact on the arrest survival times.

We now turn to our models where arrest for a new offense or supervision revocation was the outcome measure. The coefficients appear in Tables 38 and 39. Consistent with previous literature, a history of prior commitments (EPRIORCM) was significant for both men and women. Individuals who had a prior commitment were more likely to be arrested or revoked.

Age at first commitment (AGEFIRCO) and age at release (AGERLSE) had the expected effect for men. The older a man was at release and the older a man was when he was first committed the less likely he was to recidivate. These age effects were not found for women.

One of our predictors measured in-prison behavior prior to release – having a more serious – level 100 or 200 – disciplinary infraction during the 6 months before release (ERELIRY). We found that for both men and women, having one or more serious disciplinary infractions before release was related to a greater likelihood of recidivism.

The effects of drug use history and employment history covariates differed between men and women. Men who used two or more illicit drugs as well as alcohol on a daily basis in the year before their most recent arrest (TWOALCY) were more likely to be arrested or revoked than men who did not use any drugs on a daily basis during that time period. Among women, no type of drug use pattern was significant. In addition, while men who had been employed in the month before their most recent incarceration (EWORKJOB) were less likely to be arrested or revoked within 3 years after release, no effect of employment history was found for women.

The effects of post-release marital status also differed between men and women. Men who lived with a spouse (ESPOUSE) upon release were less likely to fail whereas men who lived with a common-law partner (ECOM_LAW) were more likely to fail than men on average. In contrast, no effects of post-release living situation were found among women.

Summary

In summary, we found that in-prison treatment decreased the likelihood of failure among men when failure was defined as either arrest for a new offense *or* arrest for new offense or revocation for violation of a condition of supervision. This effect was found regardless of whether or not we used a model which controlled for selection bias. We found, however, a larger treatment effect when controlling for selection bias. This was a result of the sign of the coefficient for the correlation between error terms, indicating that men who entered and completed treatment were more likely to recidivate. In contrast, we consistently found no treatment effect for women. Furthermore, we did not find evidence of selection bias among women.

Our examination of the other predictors pointed to possible gender differences in the predictors of recidivism. The only effects consistently found among both men and women were those associated with crime and serious prison misconduct. For both men and women, those with a

prior commitment or a serious disciplinary infraction within 6 months before release had shorter survival times.[15]

First Detected Drug Use as Failure

The results for failure defined as drug use, like those for arrests and revocation, were limited to those subjects who were released to supervision and for whom data were obtained from a Probation officer. In addition, we limited the analyses to subjects who had their urine tested as a condition of supervision.

We note that the intensity of urine testing decreases over time for those who successfully avoid testing positive. Of course, this means that the probability of being detected decreases over time, and consequently the estimates of the survival function for relapse to drug use conflates behavior by people under supervision (drug use) with behavior by Probation officers (monitoring for drug use). This is not a problem for our analysis provided we interpret the findings appropriately.[16] That is, judgment of "success" following release from prison comes from a combination of an objective urine test and a subjective expert judgment by a Probation officer. Approximately 54 percent of the men had evidence of drug use after release as compared to 42 percent of the women. The results for men are reported in Table 42 and the results for women in Table 43.

In looking at Table 42 we see that for men, all three models agree that substance abuse treatment was effective at reducing subsequent relapse to drug use. The parameter estimates for the treatment effect (WASTREAT) were highly significant. Moreover, the parameter estimates for the two methods that adjusted for selection bias were in substantive agreement, and both offered parameter estimates that were almost three times larger than what was derived from the unadjusted model approach that did not adjust for selection bias.

We would judge that, for men, substance abuse treatment has a demonstrably favorable effect on reducing relapse to drug use. This did not appear to be the case for women (*see* Table 43). The

---

[15]Although prior research indicates that sexually abused women do as well in substance abuse treatment as non-abused women (Gil-Rivas, 1996 and 1997; Davis, 1996), we also conducted additional analyses where we controlled for abuse history. The results of these analyses for all our outcome measures showed no effect of abuse history and also did not alter our findings regarding the effects of in-prison treatment.

[16] We note that urine testing can decrease or cease even after an individual has had a positive urinalysis test during their supervised release. However, our data does not allow us to assess the length of time elapsed between a positive test and subsequent negative tests before a change in urine testing frequency is warranted.

test of treatment effectiveness did not approach statistical significance in any of the three models. Restricting the follow-up period to 12 or 18 months did not improve the integrated hazard plot, nor did the findings change substantively. Furthermore, the treatment effect was not statistically significant in a Cox proportional hazard model.

In looking at the coefficient for the selection bias parameter – ν – we see results consistent with those we found for recidivism. The coefficient was significant for men but not for women. Furthermore, the estimated correlation between error terms in Table 40 showed that the correlation between the error terms was large for the men but not for the women. Thus, among men only, those most likely to use drugs after release, holding the observable covariates constant, were most likely to enter and complete treatment. The Heckman model thus estimated a larger treatment effect than did the unadjusted model. As with the models of recidivism, the small correlation between error terms for women, which indicated no selection bias, explained why the parameter estimates among women were similar across all three models.

In calculating the estimated probability of failure as we did with recidivism, an estimated 59 percent of untreated men would relapse to drug use compared with 50 percent of treated men (*see* Table 41). Drug relapse rates were high whether people are treated or not, but those who were treated did much better than those who were not treated. An estimated 43 percent of untreated women relapsed compared with only 35 percent of those women who were treated. Again, the differences were large, but not statistically significant.

As with the results for recidivism, we found both similarities and differences between men and women when comparing the other covariates which were found to be significant. Similar to our results for recidivism, we found a history of prior commitments (EPRIORCM) associated with a higher likelihood of failure, defined as drug use, among both men and women. We also found that for both men and women, a higher age at release (AGERLSE) was associated with a lower likelihood of drug use.

Unlike the models for recidivism, demographic characteristics were found to be related to drug use. African American men and women (EBLACK) were more likely to use drugs after release than other men or women on average. Ethnicity was related to post-release drug use only among men. Hispanic men (EHISP) were more likely to use drugs after release.

Previous drug use and drug treatment were associated with post-release drug use for men but not for women. Men who had a previous history of drug treatment (EPSTDGTX) were more likely to use drugs. In contrast, women who had a history of previous mental health treatment (EPSTMHTX) had longer survival times. Among men, all but two types of daily drug use patterns were associated with a greater likelihood of drug use when contrasted with a pattern of no daily drug or alcohol use in the year before arrest. The drug use types associated with a greater

likelihood of failure were: marijuana without alcohol (MJNOALC), one illicit drug other than marijuana without alcohol (ONEALCN), and two or more illicit drugs other than marijuana, with or without alcohol (TWOALCY, TWOALCN). Among women no drug use category was associated with a longer survival time. However, women who had never used an illicit drug on a daily basis (ENODAILY) were less likely to fail than those who did not use any drugs on a daily basis in the year before incarceration.

Unlike the analyses of recidivism where our in-prison disciplinary infraction predictor included the more serious infractions, our analysis of post-release drug use limited the infractions to those that were drug related. We note that drug infractions are a subset of the more serious infractions. The results indicated that among both men and women, a drug related infraction before release (EDRUGIRY) was associated with a higher likelihood of drug use after release.

Similar to the results for arrest or revocation, we found that living situation after release was related to post-release failure for men only. Men who lived with a spouse (ESPOUSE) were less likely to use drugs than men on average.

To summarize, we found that in-prison treatment decreased the likelihood of failure among men when failure was defined as evidence of post-release drug use. This effect was found in all models, the model which did not adjust for selection bias and the two models which did adjust for selection bias. We found, however, a larger treatment effect when we adjusted for selection bias, more so when we used the Heckman approach than when we used the instrumental variable approach. The larger effect when adjusting for selection bias was a result of the sign of the coefficient for the correlation between error terms in the Heckman model, indicating that men who entered and completed treatment were more likely to use drugs. In contrast, we consistently found no treatment effect for women and we found no evidence of selection bias among women.

Post-Release Employment

Next, we examined employment during the supervision period following release. Although the primary emphasis of the BOP's drug treatment programs is to decrease post-release drug use and criminal behavior, it also emphasizes prosocial behavior. The outcome measure was the percentage of time spent working. It was censored at 0 percent and 100 percent, so we used a two-limit tobit model to test whether or not substance abuse treatment increased employment rates. Because some individuals were not employed at all during the entire post-release period, we included a dummy variable identifying these individuals as a control variable in our analyses. On average, men were employed 69 percent of the time– during the up to 3-year follow-up period, and women were employed 62 percent of the time.

None of the three approaches suggested that treatment improved employment rates for men (*see* Table 44). Curiously, however, all three approaches suggested that being treated improved employment rates for women (*see* Table 45). This is curious because there was no evidence that treatment reduced relapse to drug use for women, so it is difficult to tell a story about how substance abuse treatment leads to improved employment outcomes without a finding that treatment also reduced drug use among women. We should note, however, that the effect was not highly significant and that while the effects of treatment on drug use were not significant they were in a positive direction, that is, treated women had lower drug use rates.

Another way to look at employment as a criterion value is to characterize work as (1) full-time during all of the follow-up period, (2) full-time during part of the follow-up period, (3) part-time, and (4) no job. We refer to this measure as employment level. Subjects who were unemployed because they were retired, disabled, homemakers, or in school full time were not included in these analyses, and this group consisted of 56 men (3.2 percent of the men released to supervision) and 25 women (5.3 percent of the women). For men, the final analysis sample size was 1,775, and for women it was 447. Thirty-seven percent of the men were employed full-time the entire post-release period as compared to 22 percent of the women. An approximately equal percentage were not employed at all after release — 10 percent of the men and 9 percent of the women. Thus a greater percentage of women were either employed full-time some of the post-release period (56 percent of the women versus 46 percent of the men) or employed part-time (13 percent of the women verus 7 percent of the men).

We analyzed this outcome using the unadjusted approach as well as the instrumental variable approach.[17] Results are summarized in Table 46. Parameter estimates from the unadjusted model were virtually identical to estimates from the instrumental variable model. There was no evidence that substance abuse treatment affected employment levels among men. However, the instrumental variable model suggested that the employment level improved for women. Given the previous finding – that the rate of employment improved for women following treatment – we might conclude that treatment improved post-release employment for females.

The coefficient representing selection bias (ν) was available only for the model of employment rate. For both men and women we found that ν was not significant. Furthermore, the correlation of the error terms was very small for both men and women (*see* Table 40). These results explain why the parameters for the treatment effect – WASTREAT – were very similar across all three modeling approaches among both men and women.

[17]Project constraints limited our ability to do Heckman models for all outcome measures.

When we calculated the employment probability for treated and untreated subjects, among men we saw no substantive difference. The employment probability was .70 for treated men and .69 for untreated men. Similarly, 36 percent of the treated men were employed full-time their entire post-release period as compared with 38 percent of the untreated men. In contrast, among women, we found substantial differences between the treated and untreated subjects. The employment probability among treated women was .69 as compared with an employment probability of .59 for untreated women. Similarly, 31 percent of the treated women were employed full-time their entire post-release period as compared with 18 percent of the untreated women (*see* Table 41).

For the most part, within each gender group, the covariates which were significant in the models of employment rate were also significant in the model of employment level. It is noteworthy that in comparing the significant predictors across gender, there were very few predictors of post-release employment for women.

Among men, several demographic predictors were significant. Hispanic men (EHISP) had lower levels of employment. Among women, we found that ethnicity was related only to employment rate and not employment level: Hispanic women had lower employment rates. Among men, individuals of other race (ERACEOTH) had lower employment rates whereas individuals with a higher level of education (GRADEA) had higher employment rates.[18] These variables were not significant for women. Men who were older at the age of first commitment (AGEFIRCO) had higher employment rates and levels, whereas men who were older upon release (AGERLSE) had lower employment rates and levels.

Previous employment history was related to post-release employment among both men and women. Men and women who were never unemployed for a period of at least 30 days (NVERUNEM) had higher employment rates and levels than subjects on average. Not surprisingly, among women, those who had never worked (NVRWORK) had lower employment rates and levels. Among men only, those who were employed in the month prior to incarceration (EWORKJOB) had higher levels of post-release employment. Among both men and women, those who were not in the work force before incarceration (ELEGITUN) had lower post-release employment levels.

Drug use was related to post-release employment among men but not women. Men who used one illicit drug (other than marijuana) but not alcohol on a daily basis in the year before arrest (ONEALCN) had lower employment rates and levels than those who had no daily drug use.

---

[18] We note that ERACEOTH was not significant in the model of employment level for men but that it approached significance.

Furthermore, men who used two or more illicit drugs (other than marijuana) and alcohol on a daily basis (TWOALCY) had higher employment rates.[19]

Similar to the results reported in our models of recidivism and drug use, living situation was related to employment after release, but only among men. Men who lived with their spouse upon release (ESPOUSE) were more likely to be employed full-time and employed a greater percentage of the time. In contrast, men who lived with a paramour (ECOM_LAW) had lower employment rates. However, this variable was not significant in the employment level models. In summary, we found effects for DAP treatment on employment only for women. We also found no evidence of selection bias for either men or women.

Community Corrections Center (CCC) Placement Failure

During a Community Corrections Center placement, an individual is ordinarily housed in a halfway house but may also spend the latter part of this placement on home confinement. CCC placements are governed by policy (which excludes some individuals from receiving such a placement), and by the discretionary decisions of wardens. Seventy-three percent of the men in our sample and 77 percent of the women received such a placement. Twenty-three percent of the men and seventeen percent of the women failed at a halfway house placement. Our results for halfway house placement are limited to the unadjusted model and instrumental variable approaches (*see* Table 47). The unadjusted model would cause us to conclude that substance abuse treatment had a strong and beneficial effect on CCC placement outcomes for men. The t-score is of unquestionable significance at 4.51. After adjusting for selection bias, however, the results are much different and give us no reason to believe that substance abuse treatment improved the CCC experience. If only the unadjusted model were considered, we might mistakenly conclude that substance abuse treatment helped offenders succeed in this transitional period between prison and community supervision. For women, there was no disagreement between the two models, in that substance abuse treatment seemed to have no statistically significant effect on the CCC placement experience.

The criminal history covariates that were significant in the CCC placement models were similar for men and women. Among both men and women, individuals who were first incarcerated at an older age (AGEFIRCO) were less likely to fail their CCC placement. In contrast, individuals who had a prior commitment (EPRIORCM) were more likely to fail this placement. Although having been employed in the month before arrest (EWORKJOB) was related to CCC placement failure, the direction of this coefficient differed between men and women. Men who were employed were more likely to fail their CCC placement whereas women who were employed were less likely to

[19] This coefficient was near significance in the model of employment level.

fail their CCC placement. Two other covariates were found to be significant only among men. Men who were not in the work force before their most recent arrest (ELEGITUN) were less likely to fail their CCC placement. In addition, men who had lifetime diagnoses of both antisocial personality and depression (EDIAGBTH) were more likely to fail their CCC placement.

We note that the variables indicating whether or not the individual had up to a one-year sentence reduction available (EYEARYES) were not significant for either men or women. This suggests that individuals who had already successfully completed their in-prison treatment phase were not affected by the possible loss of their sentence reduction were they not to successfully complete their CCC placement.

Summary of Results

The findings we reported represent a coherent pattern. Substance abuse treatment reduced relapse to drug use for male offenders. A lower rate of substance use after release apparently leads to a lower level of criminal offending. We have not explicitly tested this assertion in a causal setting, but it is a reasonable inference to draw from these findings. Furthermore, treatment did not seem to reduce relapse rates for women (although treated women had lower relapse rates), and if drug use "explains" criminal recidivism, then we would not expect treatment to affect criminal recidivism for women. Evidence is consistent with that conclusion.

The following provides a summary of the effects of treatment for each outcome measure, and also highlights those covariates which were consistently found to be related to these outcomes.

Recidivism: Arrests and/or Revocations

We found positive effects of in-prison residential drug treatment on our various indicators of recidivism. Those indicators were arrest for a new offense among supervised subjects, arrest for a new offense among both supervised and unsupervised subjects, and arrest for a new offense or supervision revocation among supervised subjects. The treatment variable was a probability representing the conditional probability of completing treatment given that one entered treatment. This variable should not be confused with following only treatment completers as has been done in previous research.We have argued that by assessing only treatment completers, the analysts may not be able to separate the effects of treatment from the effects of self-selection. The conditional treatment completers were less likely to recidivate than untreated men. Among women, although treated subjects were less likely to recidivate than untreated subjects, the results were not statistically significant.

We also found gender differences when assessing whether there was evidence of selection bias and if there was such bias, the nature of the bias. Men who entered and completed treatment were more likely to fail, indicating that without adjusting for selection bias, the effects of treatment were weaker than when adjustments for selection bias were made. Women, on the other hand, did not show evidence of selection bias.

Only two covariates were found to be significant among both men and women. Individuals who had a prior commitment and individuals who had a serious disciplinary infraction in the 6 months before release were more likely to recidivate.

Drug Use

Our results for drug use were similar to those for recidivism. We found significant effects of treatment for men but not for women. We also found evidence of selection bias among men but not women. Treated men were less likely to use drugs after release than untreated men. In addition, treated women were less likely to use drugs after release than untreated women, but the results were not statistically significant. Among men, those likely to use drugs during the follow-up period were more likely to enter and complete treatment. Once again, among men, the models that adjusted for selection bias tended to estimate a stronger treatment effect than those models that did not adjust for selection bias.

Men and women shared several predictors of failure, when failure was defined as drug use. Men and women who had a prior commitment, were African American or had a drug-related disciplinary infraction within the 6 months before release were more likely to use drugs after release. On the other hand, men and women who were older upon release were less likely to use drugs.

Employment

The results for employment were similar to the previous models in that there were gender differences in the effects of in-prison treatment. However, unlike our previous measures where we found a treatment effect for men but not women, we found a treatment effect for women but not for men. Treated women had higher employment rates and treated women were more likely to have been employed full-time the entire post-release period than untreated women. However, among men we found no difference between the treated and untreated subjects.

We did not find evidence of selection bias for either men or women in our model of employment rates, our measure of employment which was modeled using the Heckman approach.

Once again, we found some predictor covariates shared by men and women. Not surprisingly, men and women who were never in the work force had lower employment levels and a lower employment rate after release. In contrast, men and women who had never been unemployed for 30 or more days prior to incarceration had higher employment rates and employment levels after release.

Community Corrections Center (CCC) Placement Failure

Our outcome variable of CCC placement failure was the only outcome measure for which we did not find positive effects of having participated in in-prison residential drug treatment. The discussion of selection bias for these models was somewhat limited because we used only one method to address selection bias, the instrumental model. Thus, we were not able to assess the nature of selection bias, if any existed. Nonetheless, we noted that in the model for men, treated individuals were less likely to fail their CCC placement than untreated men in the unadjusted model. However, when we adjusted for selection bias using the instrumental variable method, this effect was no longer present. These findings suggest that, for men, without adjusting for selection bias, we might have erroneously concluded that treatment was related to CCC placement failure when in fact it was not related.