CLASSIFYING INMATES FOR STRATEGIC PROGRAMMING

Report prepared for the National Institute of Justice,
Grant # 98-CE-VX-0010

Eileen Sullivan, Ph.D., Principal Investigator
Carmen Cirincione, Ph.D., Research Consultant
Katherine Nelson, Ph.D., Research Consultant
Jennifer Wallis, M.A., Research Associate

Vera Institute of Justice
March 2001

This project was supported by Grant No. 98-CE-VX-0010 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

We gratefully acknowledge the contribution made to this report by Dr. Michelle Lang, and also by Douglas Young, who played a key role in developing and managing the project as a Senior Research Associate at Vera. We are also indebted to the staff of the NYC Department of Correction for their work and advice throughout the project.

Additional copies of this report can be obtained from:
Communications Department
Vera Institute of Justice
233 Broadway, 12th Floor
New York, New York 10279
(212) 334-1300
FAX (212) 941-9407

Requests for additional information about the research described in this report should be directed to Eileen Sullivan, Research Director, at the above address or at email:esullivan@vera.org.

Executive Summary

Drug-involved offenders now comprise the single largest group of people in U.S. prisons and jails. In response to their growing numbers, correctional facilities nationwide have developed drug treatment programs that have proven effective in reducing substance abuse and recidivism among participants. But while many long-term therapeutic programs exist in state and federal prisons, substantially less treatment is available in jails. Large cities like New York have begun to implement jail-based treatment programs but face the challenges of trying to serve a group of people whose stays in jail are temporary and unpredictable. After only a few days or weeks, inmates may be transferred to prison or make bail and return home. The erratic movement of people in and out of jail makes it difficult for corrections officials to identify inmates who are likely to remain long enough to benefit from a treatment program.

In an effort to improve its methods of selecting jail inmates appropriate for drug treatment, the New York City Department of Corrections (DOC) asked researchers at the Vera Institute of Justice to design and test statistical models that could predict inmates' approximate length of stay. Specifically, DOC wanted to identify inmates headed either for prison or the community who could complete a treatment program of 45 days or more. Vera's staff constructed two models using logistic regression and survival analysis and applied them to two sets of data. The first set contained data from DOC on a group of inmates eligible for admission to the treatment program, including demographic characteristics and information on the current case. The second set included the same records from DOC supplemented by data from other criminal justice agencies, such as additional criminal history and community ties.

Out of a hundred new inmates entering jail, Vera's researchers could correctly predict the outcome—prison or community—for about 80 people. The analyses successfully identified prison-bound inmates about 60 percent of the time and community-bound inmates about 90 percent of the time. Once researchers had determined outcome, they found that the vast majority of prison-bound inmates—nearly 85 percent—remain in jail at least 51 days. These inmates are ideal candidates for a long-term therapeutic program, which could segue into prison-based treatment. Among inmates bound for the community, the researchers were able to predict which people were likely to stay in jail for at least 37 days and who could gain more from a concentrated program helping them connect with resources in the community. Neither statistical model, however, could predict which community-bound inmates would stay for certain periods beyond 37 days and who could complete a two-month treatment program.

Despite this uncertainty, DOC officials can still effectively distinguish prison and community-bound inmates and match them with the most appropriate version of its drug treatment programs. To offset any imprecision surrounding the length of stay for particular community-bound inmates, the researchers suggest that DOC modify the structure of its treatment curriculum. In the beginning the classes could focus on

managing the transition to the community, saving the more time-consuming and therapeutic segments until later in the program for those who could benefit most. Overall, jail-based treatment programs can adapt to their changeable environment by creating self-contained curricular modules that can stand independently.

Contents

Chapter One

I. Introduction

Background

Typical of many states, New York has witnessed an increase in the number of people attending its prison-based programs devoted to treating inmates' substance abuse problems (New York State Department of Correctional Services, 1999). This growth has been spurred by state and federal support for intensive, long-term programs that treat inmates for a minimum of six months but usually for a year or more (Lipton, 1995; New York State Department of Correctional Services, 1999). Many urban jail systems have made similar efforts to respond to swelling populations of inmates with drug problems. New York City now runs one of the largest jail-based programs in the country, with 1,548 beds administered by the Substance Abuse Intervention Division (SAID) of the Department of Corrections (DOC).

Corrections administrators and researchers alike, however, have acknowledged that jail settings pose uniquely challenging environments for mounting treatment programs (Tunis et al., 1996; Peters, 1993). The emphasis on movement in jails—and particularly the dynamic ebb and flow of inmates housed in pretrial detention for short and unpredictable stays—makes it difficult to design and implement systematic and efficient programs for this population. At Rikers Island, inmates can begin SAID treatment but suddenly leave the program after just a few days because they make bail or have their case dismissed. DOC also needs to determine if it should invest eight or nine months of treatment in an inmate who may end up transferring to state prison for several years.

These scenarios and the larger problem of designing jail-based treatment prompted a collaborative effort by DOC and researchers from the Vera Institute to better target SAID treatment. This paper presents the results of those efforts, which focused on early identification of inmates who could benefit from different DOC treatment regimens. Researchers developed statistical models on data that could be applied to inmates on the eleventh day after admission to jail—the same day they are screened for admission to SAID treatment—classifying inmates into four groups: people who are likely to be released to the community and those who will be directly transferred to prison, and within these two groups, people likely to stay in DOC custody long enough to complete program modules designed to prepare them for continued treatment in community or the state prison system.

Drug Treatment in Prisons and Jails

Prison-based Treatment Programs. It is widely agreed that the growth in jail and prison commitments over the past two decades is largely attributable to escalating numbers of drug-involved offenders. As the rate of incarceration increased about 200% between 1980 and 1985, the proportion of drug offenders in jails and prisons roughly tripled, to the point that this group now accounts for the largest single group of offenders under custody (CASA, 1998; Peters and Steinberg, 2000). In a recent national survey of inmates in state and federal prisons, 83% reported some past drug use and 57% reported use in the month prior to their offense (Mumola, 1999).

The treatment response to these changes in the inmate population began to emerge in the late 1980s and early ’90s, particularly in the form of prison-based treatment programs. Two federal initiatives, Project REFORM, funded by the Bureau of Justice Assistance, and the later Project RECOVERY, funded by the Center for Substance Abuse Treatment, provided technical assistance to states in support of their efforts to develop and implement programs. More recently, Congress established the Residential Substance Abuse Treatment (RSAT) Program within the federal Department of Justice to invest $270 million in state and local corrections treatment. RSAT funds are restricted to programs that treat inmates for a minimum of six months, so, like the previous REFORM and RECOVERY efforts, the RSAT initiative focuses much more on prison- than jail-based programs. Like many other states, New York’s Department of Correctional Services developed and implemented the CASAT (Comprehensive Alcoholism and Substance Abuse Treatment) program in 1990. Over 1,500 inmates are in CASAT at any one time, and the program has served over 26,000 inmates since October 1990 (New York State Department of Correctional Services, 1999). The Department also operates a number of less intensive ASAT programs and in the past year opened new programs with federal RSAT funding.

Early studies of another New York prison program, Stay ’N Out, and programs in Oregon and Delaware suggested they were effective in reducing recidivism among program graduates (Wexler et al., 1988; Field, 1989; Inciardi et al., 1997). These studies received wide attention from researchers and administrators who supported prison treatment. While the methods employed in the early studies have been subjected to considerable criticism (Gaes, 1999), a number of more recent studies, including some with improved designs, have continued to support these programs. Outcomes consistently indicate that inmates who stay in programs for a minimum of three months, and particularly those who complete post-prison aftercare treatment, show better outcomes than comparison groups who do not attend treatment or drop out early (Martin et al., 1999; Knight et al., 1999; Wexler et al., 1999).

The accumulated evidence in support of these programs has prompted several reviewers to promote policies integrating drug treatment in correctional settings and to argue more specifically for expanding programs for inmates sentenced to prison and released to the community (e.g., Lipton, 1995; CASA, 1998). Besides the evident service needs of the population, central to the underlying logic of these programs is that inmates present a stable, “captive audience” who can be exposed to the requisite period of treatment. Inmates in prison also have little competing stimuli and are usually responsive to pressure from legal contingencies, such as the promise of early release for entering and completing a treatment program.

Jail-based Programs. Support for jail-based treatment programs has been less plentiful in both the policy and research literature. This probably is due in part to the fact that there is less treatment offered in jails. A recent national survey by the Justice Department's Bureau of Justice Statistics (BJS) indicated that 14% of jail inmates who reported regular use of drugs or alcohol said they had attended a substance abuse treatment program in jail after admission (Harlow, 1998). Another national jail survey reported that eight percent of inmates participated in drug treatment programs (CASA, 1998). In contrast, BJS's most recent survey of state and federal prisoners indicated that 36% of inmates reporting regular use of drug or alcohol attended a treatment program after admission to prison (Mumola, 1999).

The reduced prevalence of programs in jails cannot be attributed to the less urgent needs of inmates. In the nationwide federal studies, 60% of jail inmates reported using alcohol or drugs at the time they committed the current offense and it is estimated that 27% had drunk enough alcohol to be impaired when the crime was committed. Fifty-five percent of surveyed jail inmates reported drug use in the month before their arrest, 24% reported using cocaine or crack, ten percent used amphetamine and other stimulants, and nine percent used heroin. These figures are five to ten percent higher than comparable figures for prison inmates (Mumola, 1999). In one recent study, BJS estimated that three-quarters of all prison inmates could be characterized as alcohol- or drug-involved offenders. Much less detailed data are available on jail inmates, but they would likely show similar results.

Even if substance abuse problems occur in similar numbers in jails and prisons, the appearance is more acute in jails. Prisons admit sentenced inmates who have usually been held in local jails for months after their arrest. In contrast, jails admit people off the street and in diverse legal circumstances, ranging from alleged offenders detained pre-trial to inmates sentenced to jail terms of less than a year. Users entering jail might still be high, in withdrawal, or craving more drugs. They may be actively supporting a family or involved in a treatment program and had a lone relapse. They may present acute medical or psychological problems that are associated with chronic drug use. Compared to prisons, jail settings lend an air of immediacy and urgency to drug problems; substance abuse is part of a constellation of public health problems—social, mental, and physical—presented by many newly admitted inmates.

This sense of urgency has led New York City DOC, along with several other large jail systems, to create some kind of treatment response to the jailed population. The easiest and most efficient way to provide treatment in jails is to design it like a prison program—exclusively for inmates admitted to serve known jail sentences of, for example, six months to a year. Some urban systems, such as Florida's Dade County (Miami) and

Harris County (Houston) in Texas, offer programs only to sentenced inmates. The SAID program operated by New York City DOC has 648 treatment beds identified for sentenced inmates.

But the balance of SAID's beds—900, or 58% of 1,548 slots—are for inmates who begin their stay in custody as a general population pre-trial detainee. National statistics indicate that detainees make up just under half of city and county jail populations (Harlow, 1998), while an informal survey done for this research suggests the figure may be closer to two-thirds in large, urban jurisdictions.[1] The short stays of inmates sentenced to 30, 60, or 90 days make jail populations dynamic, difficult targets for service programs. It is the detainee population (and inmates suddenly leaving with sentences of time served), however, that bring the most frustration to administrators and staff who see the treatment needs of this large and important group of inmates. In two of the five jail programs studied by Tunis and colleagues (1996), nearly two-thirds of participants did not complete the treatment curriculum because they were released by the court. In addition to the sudden and unpredictable departures of these inmates (some of whom go to prison), detainees may also move frequently while under custody for court appearances, meetings with attorneys and police, and mandated medical and mental assessment and care. For these reasons, community-based agencies are reluctant to invest staff time working with detainees to prepare them for discharge to the community and to help them link to services upon release.

As the treatment funnel gets smaller and smaller, it is not surprising that one survey showed that detainees were about half as likely to obtain treatment as sentenced jail inmates.[2] Some jail systems do make efforts to treat this group. About one-quarter of the 300 participants in Philadelphia's jail-based program are detainees, while local corrections staff in Cook County (Chicago) reported to us that about 85% of their treatment participants are detainees. To cope with the uncertain status of detainees, the Cook County program uses what might be termed a clinical approach to selecting participants—employing staff experienced with court decision making to identify and recruit detainees who are likely to stay for a sufficient duration to benefit from the treatment.

In discussions leading to this research, New York City DOC administrators acknowledged they had no systematic means of identifying inmates who were likely to stay for requisite periods, and expressed considerable interest in having some basis for doing so. Plans for developing a statistical tool for predicting length of custody for SAID-appropriate inmates were prompted both by DOC's interest and by past experience at the

[1] Sixty-four percent of New York City DOC's population are pre-trial detainees. Through telephone calls to local departments of correction we learned that 72% of Miami's inmates are detainees, 67% of Philadelphia's inmates are detainees, and 62% of Houston's inmates are detainees.

[2] Harlow (1999) recounts that 14% of sentenced jail inmates report attending a treatment program after admission while 7% of detainees report attending a program.

Vera Institute in developing predictive models of sentences. The objectives and methods of the study were responsive to that collaboration.

Objectives of the Study

Our main goal was to construct and test a statistically based system of classifying inmates for one or more types of SAID-sponsored substance abuse treatment programs within the DOC system. A number of more specific objectives regarding construction and analysis (such as gathering and cleaning databases from multiple sources and testing multiple analytic schemes) followed from this goal.

As the work progressed, more immediate goals emerged from ongoing discussions with DOC administrators. On the tenth or eleventh day after admission to the system, detainees are screened and identified for SAID treatment (sentenced inmates receive SAID treatment at separate facilities). DOC planned to create two sets of treatment services for detainees entering treatment at that point—one for those who are likely to go to state prison and another for those who are destined for release to the community after serving a jail sentence or serving time for pre-trial detention. Each treatment component required about 45 days to administer. While inmates staying more than 45 days could receive progressively more advanced programming, DOC's primary objective was to be able to identify two groups on the day of SAID screening: detainees who would be released to the community but would stay approximately 45 or more days in the system, and detainees who would go to state prison but remain at least 45 days in the system. DOC was making plans for the city's Human Resources Administration to provide services that would enable released inmates to obtain entitlements upon discharge to the community. These services required that inmates stay at least 45 days after the entitlement paperwork was initiated; if DOC could identify detainees who were released to the community but who stayed approximately 45 days or more, these inmates would also benefit from the HRA service. Building statistical models to identify these two groups became the central objective of the study.

II. Analytic Approach and Method

Statistical prediction models have been used in criminal justice research for a variety of purposes (Gottfredson, 1987; Gottfredson & Tonry, 1987). They offer insight into future events, behaviors or states, based on some statistical manipulation of information available from the past and tested in the present. Much of the research in this area has focused on the prediction of offenders' future behavior, such as appearing for trial, committing new crimes, failing on probation or parole, and escaping from an institution (Morris & Miller, 1985; Lattimore & Linster, 1996; Sims & Jones, 1997).

Predictive studies have also considered decisions by various actors, such as judges, police, victims, and correctional officials, at different points in the criminal justice system. Some of these studies have sought to predict type of case disposition and length of custodial sanctions for different types of offenders. Predictors, or characteristics that indicate certain outcomes, identified in this literature have informed policy and have been used by criminal justice officials to classify offenders and make decisions about sentences, release, parole, probation supervision, and treatment (Glaser, 1987; Hepburn, 1994). The value of this research is evident when the predictions are based on statistically robust methods (Clear, 1998).

Studies that come closest to addressing the kinds of classification problems posed by DOC's SAID administrators are those that have predicted sentence outcome, and more specifically incarceration and length of sentence. While these studies have used different statistical approaches, they have employed fairly consistent predictors. Core variables include severity and type of the immediate charge, prior criminal record, and personal characteristics such as gender, race and ethnicity, and employment status. Other predictor information, less readily available, include additional legal measures (such as gun involvement, type of crime, and presence of open warrants) and case-processing factors (attorney type, bail amount set, and pre-trial release status).

Using the subset of these variables available to us, we took an analytic approach similar to that used by researchers from Vera and New York City Criminal Justice Agency (CJA) in studies conducted in the early 1990s (Winterfield, 1992; Belenko et al., 1994, 1995). These studies were focused on more narrow and predictable sentencing outcomes but, like the current research, they were also based on data from the DOC Inmate Information System (IIS) and CJA. Researchers used logistic and ordinary least squares (OLS) regression models to predict the likelihood of receiving an incarcerative sanction (jail or prison) and the likely length of that sentence**.** Predictors consistently found to be significantly related to disposition and length of sentence included the severity of the indictment charge, the number of felony arrests, and the presence of one or more prior felony convictions. The goal of that research was to estimate the displacement effects of various intermediate sanctions on jail and prison beds in New York City. In the present study, our goal is to examine the predictive capacity of these and other variables on length of stay in jail using logistic and survival analyses.

Samples and Data Sources

DOC Data. The analytic sample used for the prediction models was a sample of recent jail inmates identified as eligible for admission to SAID. Selected from the DOC IIS database, the initial sample included 29,716 people who entered DOC custody as detainees (the sample did not include offenders entering with jail sentences) and who were released between July 1, 1997, and June 30, 1998.[3] This group was then screened for SAID eligibility. Inmates could not have higher scores than 16 on a DOC security classification measure (generally meaning that they not have a history of institutional violence or escape).[4] Additionally, to be eligible for SAID, inmates could not have any of the following: convictions for any weapons-related infractions, convictions for any violence-related infractions within five years, arson-related convictions or charges, or convictions or current charges for sex-related offenses. Further screening criteria were specified through discussions with DOC to ensure that the data set included only detainees who would be eligible for the new, planned SAID (and HRA) programs. This meant ensuring that detainees in the sample had been in custody at least eleven days (and had not received a jail sentence by that time) and were not technical parole violators.

We also excluded from the sample small subgroups of inmates identified as cases who would be treated differently by the court (or appeared to have been based on their actual time in custody) when compared to the general population of SAID-eligible inmates. These cases totaled about 1,500 and included inmates who had absconded or escaped from DOC custody; left custody as a result of death, suicide, a warrant being lifted, or transfer to other jurisdictions of the criminal justice system (such as the Division for Youth, the INS, the New York State Police, or the U.S. Marshalls and other officers), or left custody for an unknown reason. The rest of the cases were excluded on the basis of SAID screening, resulting in a sample of 23,875 cases for the first set of analyses. After preliminary analysis, another 593 cases were excluded because of missing data, resulting in an analytic dataset of 23,282 cases. DOC provided about 55 data items from its IIS database on each of the cases. Apart from the inmate's entry and release dates and release status (e.g., made bail or released on recognizance as detainee, completed jail sentence, transferred to state prison), these data fell into three main categories: demographic characteristics (including gender, ethnicity, age, self-reported drug use, marital status, and citizenship; information on the current case (such as bail amount, arraignment charge and severity of charge (charge class), whether the charge was for a violent crime, whether the person was remanded, date of arrest, borough of arraignment, and type of warrant (if any) existing at time of arrest); and a few simple indicators of prior criminal record (items from DOC's classification instrument, including histories of prior convictions, escape, violence, institutional conduct in the past five years).

While original plans called for us to gather additional publicly available data on each case for purposes of analyses, DOC administrators indicated their interest in conducting one set of analyses that used DOC data only. They expressed pragmatic concerns about the department's ability to quickly routinize the process of obtaining and integrating the

[3] The release could be for any number of reasons, such as making bail, having charges dismissed, completing a contiguous jail sentence, or transferring to state prison.
[4] Those cases in the data set with classification scores of over 16 are due only to a recalculation of classification score in order to address missing data and include as many cases as possible in the analyses.

added data so that it could be employed in classifying inmates for the new SAID program within 11 days of admission to custody. We agreed to conduct two sets of analyses: one that employed only DOC data and another that included the supplementary data from DCJS and CJA (which would presumably yield more accurate prediction models).

Adding DCJS and CJA Data. Much more complete information on all three data categories is available from the state's repository of criminal justice data, maintained by the Division of Criminal Justice Services (DCJS). Some additional DCJS data items include arrest charges; Uniform Crime Report charge categories; and flags indicating whether the charge is for an offense involving drugs, a firearm, another weapon, or a child victim, a DWI or DWAI arrest, or whether the person is eligible for probation We also sought three other data items from the New York City Criminal Justice Agency (CJA) that were not available from other sources and might have some predictive value. Recorded from interviews done by CJA to make recommendations regarding pretrial release, these items included length of time a defendant has been residing at current address; whether the defendant is employed, in school or in a training program full-time; and CJA's release (ROR) recommendation.[5]

While matching and merging the DCJS and CJA data with the original DOC sample yielded many more data items, it inevitably reduced the size of the analytic data set and introduced bias into this smaller sample. To obtain the DCJS data, we provided them with state identification numbers (NYSIDs) and arrest dates for the original 29,716 cases. DCJS returned perfect matches—the same identification number and arrest date—on 20,927 cases. Of these initial matches, 5,017 of the current cases were sealed (because the charge was dismissed or it was a disposed as a juvenile case) and could not be released to us under state statute, resulting in information on 15,910 cases. Although the cases for which DCJS had no match were likely a random group (with miscoded identification numbers or arrest dates), the exclusion of the sealed data meant that lower-level cases (those resulting in dismissals) were underrepresented in the sample. At this point, data from CJA was also merged into the existing data set, which did not affect the sample size. Since the information returned by DCJS included multiple events when they existed for a given arrest date (creating multiple lines of data for some of the cases), the data set was reduced to information associated with top charge for each arrest date. This ensured that the DCJS data would be comparable to the DOC data, also based on top charge, resulting in a sample size of 15,438.

The additional SAID eligibility screening that was applied to the DOC data set was then applied to the DOC/DCJS/CJA data set. Again, detainees who received a sentence

[5] CJA's release recommendation, which measures the strength of the defendant's ties to the community, is used in making a recommendation to the arraigning judge as to whether the defendant would have a lower risk of failure to appear if released (Belenko, 1993).

within ten days of admission to DOC were excluded from the sample, as were subgroups of cases that were different from the general population SAID-eligible case. The resulting sample of the merged, multisource data set included 12,912 cases. After preliminary analysis, another 163 cases were excluded because of missing data, resulting in an analytic dataset of 12,749 cases.

Analytic Plan

Analyses were conducted in four phases. Initially, we performed simple descriptive analyses to examine the key dependent variables in the research. For example, we created plots of time-in-custody for the two major groups—inmates released to the community (as releasees or as those completing jail sentences) and those transferred to state prison—and discussed them with DOC researchers and administrators. We also examined frequency distributions of time-in-custody for other subgroups of interest (people released pretrial, or those released after completing a jail sentence served contiguously with pretrial detention). We assessed the various distributions separately for both the larger DOC sample and the DOC/DCJS/CJA sample.

The second, data reduction phase, involved several discrete analyses. We conducted bivariate statistical tests (correlations, t-tests, and chi-squares) to assess multicollinearity among the predictor variables and to select variables that were shown to covary with each dependent measure. Variables of theoretical interest (because of their hypothesized association with the outcome) were also kept in the analyses. Again, we performed separate analyses on the two databases for the prison-bound and the releasee groups. We also ran analyses to resolve missing data on variables that were of interest; for example, with continuous variables (e.g., bail amount), cases with missing data were assigned the mean, median, or mode of the subgroup to which they belonged (e.g., males with no prior convictions charged with a B felony).

The final analysis in phase two investigated ideal segmentation of time-in-custody. Although DOC had identified 55 days as the ideal cutoff point (since inmates were required to stay about 45 days after being identified for the treatment on the tenth or eleventh day of custody, the length of stay count would start at day eleven, not day one), the strength of any statistical model aimed at distinguishing groups on either side of a given cutoff point would vary depending upon the actual distribution of the time-in-custody variable and the available predictors. It might be that inmates who leave before day 30 are consistently different from those who leave after day 30 and thus these two groups can be identified statistically, while it is very difficult to distinguish groups of inmates who leave before and after 75 days (yielding weak and inaccurate statistical prediction models).

To assess the "natural" cutoff points, we examined the distributions of the time-in-custody variables with DOC researchers and administrators while also producing Chi-Squared Automatic Interaction Detector (CHAID) algorithms. Developed by Kass

(1980), CHAID can evaluate all the values of a variable (in this case, days in custody) to identify those most associated with a criterion variable (prison versus community release). Using as a criterion the significance of a statistical test (Chi-Square for continuous variables and Likelihood Ratio Test for ordinal and discretized continuous variables), CHAID could merge values of the time-in-custody variable that are judged to be statistically homogenous and maintain all other values that are heterogeneous. The result of this statistical process is “natural” cutoff points of length of stay by which to examine and distinguish among the groups of inmates and generate the strongest prediction models possible.

For the DOC-only data set, the natural cutoff point for length of stay identified by CHAID, on which we were to base our logistic regressions, were 37 and 51 days. By DOC standards, these points were 27 and 41 days from SAID assessment. For the combined DOC/DCJS/CJA data set, the cutoff points occurred at 45 and 61 days. Again, by DOC standards, they occurred at 35 and 51 days from SAID assessment. (Please refer to Appendix A for details on the CHAID analyses.)

Phases three and four of the analyses involved the actual multivariate model-building process. On each of the two data sets, we first constructed logistic regression models to examine main effects to distinguish the prison- and community-bound groups. Next, we separated the samples into those who were transferred to prison from DOC custody and those who were released to the community (as a releasee or as someone completing a jail sentence). We then ran another set of logistic regressions on each of these samples, examining two sets of cutoff points for length of stay identified in phase two of the analysis.

Phase three generated ten logistic regression models.[6] Predicting length of stay, first by using only information provided by DOC addresses a feasibility issue. Using only information that DOC has readily available enables the department to make predictions in order to apply appropriate treatment programs to inmates quickly and more easily than if DOC had to rely on getting information from other sources in order to predict length of stay. Logistic regressions were later run on the combined DOC/DCJS/CJA data sets, incorporating information from these public sources; these analyses examine whether predictive power can be increased by including more information from other sources.

Using the DOC-only database, five models were constructed. The first: (1) predicted membership in the prison group versus the releasee group. Among people who would go to prison, the next models predicted, based upon the natural cutoff points for length of stay identified by CHAID, (2) who would stay at least 37 days, and (3) who would stay at least 51 days. Among those who would be released to the community, the models predicted (4) who would stay at least 37 days, and (5) who would stay at least 51 days. [7]

---

[6] All ten of the logistic regression analyses use backward stepwise elimination.
[7] Throughout the body of the report, we use actual number of days from the date of admission; the days available for SAID programming after the assessment would be 10 days less than these.

Using the DOC/DCJS/CJA database, five models were similarly constructed. The first: (1) predicted membership in the prison group versus the releasee group. Among people who would go to prison, the models predicted, based upon the natural cutoff points for length of stay identified by CHAID for this data set, (2) who would stay at least 45 days, and (3) who would stay at least 61 days. Among people who would be released to the community, the models predicted (4) who would stay at least 45 days, and (5) who would stay at least 61 days.

Phase four of the analysis used another multivariate technique, survival analysis (also known as hazard modeling and event history analysis), to predict time-in-custody as a continuous variable. We performed this second set of multivariate analyses on the DOC data set to serve as a validity check on the logistic regression models, assessing whether similar variables emerged as predictors in both sets of analyses. The sample contained in this DOC-only data set includes only people who have been released from DOC custody, so it might be assumed that the use of OLS regression would be acceptable because of the lack of right-hand censoring. In models of survival data, however, assumptions about the nature of the error term (e.g., homoskedasticity and normal distribution) that are required for significance tests for the use of OLS regression are often not met. Further, the fact that only data about those who spend at least 11 days in custody will be sampled means there is effective right-hand censoring that needs to be taken into account.

Survival techniques are also advantageous to this task for other reasons. Unlike logistic models, which are most commonly used with dichotomous dependent measures, survival analysis is designed for use with continuous measures and can generate probability estimates for each value of the criterion measure for individual cases in the analysis. For any given case, then, DOC would know the estimated probability of the individual staying a given number of days. Initially, DOC might use the new treatment program on people with at least a 90 % probability of staying 55 days or more in custody. Survival models also give DOC much more flexibility if it does decide to integrate treatment components less, or more, than 45 days in length, requiring the department to predict lengths of stay other than the cutoff points we have identified. In theory (if the models are sufficiently accurate), DOC could identify inmates with high probabilities of staying in custody 180 days or more for a new six-month treatment program. In this report, however, to determine classification success of our survival models, we examined two specific cut-off points: 45 and 55 days. By DOC standards, this meant 35 and 45 days from SAID assessment.

Chapter Two

Logistic Regression Analysis of Outcome And Length of Stay at Rikers

Section I of this chapter discusses selected analysis issues, including a brief overview of logistic regression analysis and concerns that flow from the distribution of cases among the outcome and length-of-stay subgroups. Section II gauges the strength and potential usefulness of the various models that employ DOC data exclusively or the combined data from DOC, DCJS, and CJA to predict outcome—prison or release to the community—and length of stay for each outcome group. Finally, Section III outlines the variables that were significant in predicting prison or release to the community and length of stay for each of these outcome groups.

I. Selected Analysis Issues

Distribution of Cases among Outcome and Length-of-Stay Subgroups

This chapter employs logistic regression analysis (LRA) to identify predictors of outcome (prison or release to the community) for Rikers inmates, as well as characteristics that predict length of stay for each outcome subgroup.[8] LRA relates one or more predictor variables, or characteristics, to a dichotomous dependent variable, such as the outcome of prison versus release. Specifically, LRA analyzes the logit—the natural logarithm of the odds of going to prison versus being released to the community. In subsequent analyses, the logit is the natural logarithm of the odds that prison-bound or released inmates will experience one length of stay at Rikers rather than another. An odds ratio (OR) of 1.0 means that there is no relationship between a predictor and the outcome. ORs higher than 1.0 indicate a positive relationship between two variables, whereas ORs lower than 1.0 indicate a negative or inverse relationship. For example, an OR of 1.39 describes the relationship between inmates' gender and the likelihood of going to prison when men are coded 1 and women are coded 0. The OR of 1.39 means, in effect, that when other predictors in the logistic regression equation are held constant, being a man increases the odds of going to prison.

---

[8] For an overview of assumptions, concepts, and interpretation in LRA, see James T. Austin, Robert A. Uaffee and Dennis E. Hinkle, "Logistic Regression for Research in Higher Education," in *Higher Education: Handbook of Theory and Research,* John C. Smart, ed. (New York: Agathon Press, 1992), 379-410. A more technical discussion appears in Jon H. Aldrich and Forrest D. Nelson, *Linear Probability, Logit, and Probit Models* (Newbury Park, Calif.: Sage Publications, 1984), esp. 30-65. For particularly accessible discussions of probabilities, odds, and odds ratios in logistic regression analysis, see Anthony Walsh, "Teaching Understanding and Interpretation of Logit Regression," *Teaching Sociology* 15 (April 1987): 178-83; S. Phillip Morgan and Jay D. Teachman, "Logistic Regression: Description, Examples, and Comparisons," *Journal of Marriage and the Family* 50 (November 1988): 929-36.

Table 1 presents the distribution of the cases that were selected for the outcome (prison or release to the community) and length-of-stay (LOS) groups for the data set that includes information from DOC exclusively (N=23,282) and the combined data set that includes information from DOC, DCJS, and CJA (N=12,749).

The LOS categories, here and throughout the report, reflect the natural cutoff points suggested by the CHAID analysis (see Chapter One and Appendix A). In the interests of clarity, these LOS categories are in "real time." That is, they represent the number of days incarcerated at Rikers starting with the inmate's day of admission, not the number of days since assessment of eligibility for SAID services, which usually occurs ten or eleven days after admission. Readers who want to know how the predictors influence the likelihood of remaining at Rikers for a specified length of time after SAID assessment can subtract ten from the LOS designations shown. For example, inmates whose stay is 51 days or more have been at Rikers for at least 41 days following the SAID assessment.

Table 1: Length of Stay at Rikers by Outcome Group and Source of Data

	Outcome	
Length of Stay	Prison-Bound	Released to the Community
DOC Data	**% (#)**	**% (#)**
36 days or less	5.3 (382)	41.8 (6,708)
37 to 50 days	10.6 (768)	10.4 (1,665)
51 days or more	84.1 (6,079)	47.8 (7,680)
Total	100% (7,229)	100% (16,053)
Combined Data (DOC, DCJS, CJA)		
44 days or less	7.9 (357)	44.5 (3,655)
45 to 60 days	9.2 (417)	10.2 (836)
61 days or more	82.9 (3,760)	45.3 (3,724)
Total	100% (4,534)	100% (8,215)

Overall, the distribution of cases according to outcome and LOS are very similar for the DOC and Combined samples (Table 1).[9] For each LOS/outcome subgroup, the discrepancy between the two data sets is only two or three percent. In this respect, at least, the inevitable loss of cases that resulted from using multiple data sources appears not to have resulted in samples that are fundamentally different from each other.

Both data sets, however, reveal considerable differences in the LOS of prison-bound inmates by comparison to inmates who were released to the community. Using the DOC-only data as a case in point, the vast majority (84.1%) of prison-bound inmates stayed at Rikers for at least 51 days (or 41 days after the SAID assessment). Only slightly more than 5% stayed for 36 days or less. Most of the inmates released to the community, in contrast, fall at the highest and lowest ends of the LOS spectrum. Again using the DOC-only data, 41.8% of the released inmates stayed at Rikers for the shortest period of 36 days or less, and 47.8% stayed for the longest period of 51 days or more. A much smaller number of released inmates, only 10.4%, are characterized by the intermediate LOS of 37 to 50 days.

These findings have important implications. For example, the fact that the vast majority of the prison-bound stay at Rikers for at least 51 days means that information about inmate characteristics, the current case, and criminal history can contribute only modestly at best to predicting LOS for this outcome group; by itself, the fact that inmates are headed for prison very strongly suggests that they will remain at Rikers for a relatively long period of time, and long enough to participate in an intensive SAID program. Because the LOS of released inmates is much more variable, building a model to predict their LOS on the basis of background characteristics and current case information is potentially more fruitful. Throughout the analysis of LOS for the released group, however, it will be important to keep in mind that a LOS that is shorter than 51 days (using DOC-only data) usually signifies a <u>very</u> short LOS of 36 days or less (or less than one month following any SAID assessment).

II. Predicting Outcome: Prison or Release to the Community

The report reflects ten logistic regression analyses. Five analyses, predicting outcome and length of stay, included DOC data exclusively. The remaining five analyses included data from all sources combined—DOC, DCJS, and CJA—and were conducted independently of the DOC-only analyses. DOC variables entered into the equations for the combined data set were not limited to the DOC characteristics found to be statistically significant in the analyses that employed DOC data exclusively. This strategy enabled us to compare the effects of DOC variables on outcome and LOS when the other data sources are included and when they are not, enhancing our ability to identify ways in which the introduction of DCJS and CJA data modifies the inmate profiles that were generated by

[9] Detailed plots of time-in-custody for the DOC and the Combined samples are located in Appendices B and C.

DOC data alone. All ten of the logistic regression analyses used backward stepwise elimination.

The first objective of the analysis was to determine whether models could be constructed, using DOC data exclusively, to predict outcome—prison or release to the community—and then to determine whether including data from DCJS and CJA had a predictive payoff that would be worth the time and resources that this more complex analysis entails.

For the models predicting outcome, Table 2 shows the value and statistical significance of Chi-Square (Hosmer and Lemeshow Goodness-of-Fit Test), the variance explained (R^2), and the percentage of cases correctly predicted.

In predicting outcome, the DOC data produces statistically significant results and explains nearly half of the variance (48%). Using DOC data exclusively, one can correctly predict outcome 81% of the time, predictions will be in error for only 19% of the inmates. This means that for every 100 inmates assessed for SAID programming, 81 will be correctly predicted for the prison or release to the community outcomes.

As shown in Table 2, prediction accuracy varies according to the outcome group in question. Using DOC data only, nearly all predictions of release to the community will be accurate (92%). Predictions of the prison outcome will be correct less often, for 58% of the inmates. The remainder of the inmates for whom one predicts a prison outcome (42%) will end up being released to the community.

The combined dataset also explains nearly half of the variance (47%) and does a good job of correctly predicting outcome. Using the combined dataset, the model correctly predicts prison or release to the community 79% of the time, or for 79 out of every 100 inmates assessed. The combined dataset is somewhat better at predicting the prison outcome than the DOC data set. The DOC data makes the correct prediction of prison 58% of the time, but the combined data set correctly predicts prison 67% of the time. This difference, however, is modest (only 9 percentage points). Further, correctly predicting the prison outcome is the only respect in which the results of the combined data are superior to the results that one can obtain from using DOC data exclusively. In correctly predicting the outcome of community release, the DOC data are somewhat superior to the combined data (92% versus 86%).

The outcome models predict more accurately than would chance—the flip of a coin—which inmates are likely to go to prison and which are likely to be released to the community; the model using DOC-only data and the model that employs data from combined sources both make accurate predictions about outcome more than half of the time. The predictive accuracy of both models is even better than that considering that the odds of going to prison or being released are not, in fact, fifty-fifty—they are closer to thirty-seventy. Approximately one-third of all inmates (regardless of the data set used) go to prison, and approximately two-thirds are released to the community. On the basis of that information alone, one stands roughly a 33% chance of correctly predicting a prison outcome and a 67% chance of correctly predicting a release outcome even without any additional information. The fact that the DOC data correctly predict a prison outcome for 58% of the cases is a considerable improvement over the 33% accuracy one obtains from guessing. Similarly, the fact that the DOC data correctly predict a release outcome for 92% of inmates who are released to the community is a substantial improvement over the 70% accuracy that one would expect on the basis of guessing armed solely with knowledge of the proportion of all inmates who experience one outcome or the other.

As a practical matter, the “false positives” for the prison outcome (42% with DOC data and 33% with the combined data) could be less costly for program planning purposes than one might assume. As will be seen below, the profile of prison-bound inmates and those released to the community who remain in jail for any length of time (at least 37 days) are very similar. This suggests that inmates for whom one incorrectly predicts a prison outcome are similar to inmates who are actually headed for prison and, by extension, are also similar to those released to the community who stay at Rikers for at least 37 days. It is a reasonable working assumption, therefore, that most inmates who are predicted to go to prison (whether that prediction turns out to be accurate or not) will stay at Rikers long enough to participate in some SAID programming.

Table 2: Gauges of Adequacy for the Logistic Regression Models

Gauges of Adequacy	**Outcome**		**Minimum # Days at Rikers for Inmates who are . . .**			
			Prison-Bound		**Released to the Community**	
	Prison	**Comm. Release**	**Shorter LOS** ***(37/45 or >)***	**Longer LOS** *(51/61 or >)*	**Shorter LOS** ***(37/45 or >)***	**Longer LOS** ***(51/61 or >)***
Chi-Square (Hosmer & Lemeshow Goodness-of-Fit Test) **DOC Data Only**	χ^2 **=163.67, *p*<.01**		χ^2 **=14.97, *p=.06***	χ^2 **=13.10, *p=.11***	χ^2 **=35.06, *p*<.001**	χ^2 **=48.09, *p* <.001**
DOC/DCJS/CJA	χ^2 **=18.92, *p=.02***		χ^2 **=5.03, *p=.66***	χ^2 **=8.92, *p=.35***	χ^2 **=18.41, *p=.02***	χ^2 **=21.31, *p=.01***
R^2 **DOC Data Only**	**.48**		**.17**	**.17**	**.11**	**.13**
DOC/DCJS/CJA	**.47**		**.22**	**.21**	**.12**	**.15**
% Correctly Predicted **DOC Data Only**	**58%**	**92%**	**100%**	**99%**	**81%**	**60%**
DOC/DCJS/CJA	**67%**	**86%**	**100%**	**97%**	**77%**	**57%**

III. Predicting Length of Stay

The second objective of the analysis was to determine whether the models, using each data set, could predict length of stay for the prison-bound and released to the community outcome groups. Predicting length of stay (LOS) among inmates presented complications that we did not encounter in identifying predictors of outcome. The logistic regression analyses that predicted outcome identified relatively clear and plausible differences between the prison and community-bound outcome groups that were, of course, mutually exclusive. The analyses of LOS within each outcome group produced more ambiguous findings.

A main reason for this ambiguity lies in the distribution of cases among outcome/LOS subgroups discussed earlier in this chapter. The vast majority of prison-bound inmates stayed at Rikers for the longest LOS period (51 days or more using DOC-only data and 61 days or more using the combined data set), leaving little variation to be explained. In themselves, very lopsided distributions can produce unreliable estimates, regardless of the statistical technique employed.

Furthermore, the cases in the LOS categories overlap considerably. Within each outcome group, separate analyses attempted to identify factors that predict a “longer LOS” period of at least 51 or 61 days (depending on the data set used) versus all stays that were shorter than that, and a “shorter LOS” period of at least 37 or 45 days (also depending on the data set used) versus all stays that were even shorter (36 or 44 days or less). Necessarily, all of the cases that appear in the longer LOS subgroup also belong to the shorter LOS subgroup; for instance, stays at Rikers lasting at least 51 days by definition also last for the shorter period of at least 37 days. The fact that few inmates stayed at Rikers for the intermediate periods of 37 to 50 days, or 45 to 60 days using the combined data set (see Table 1), compounds the overlap.

Using DOC data to illustrate these points, of the 6,847 prison-bound inmates with a shorter LOS (at least 37 days), 89% (n=6,079) also belong to the longer LOS group—staying at least 51 days. Only 11% (n=768) belong to the shorter LOS subgroup exclusively-staying between 37 and 50 days. A similar pattern applies to inmates who were released to the community. Of the 9,345 released inmates who stayed at least 37 days (belonging to the shorter LOS group), 82% (n=7,680) also stayed at Rikers for the longer LOS period of at least 51 days. Only 18% (n=1,665) belong to the shorter LOS group exclusively, staying between 37 and 50 days.

Prison-Bound Group

The very high predictive accuracy of close to 100% of the DOC and combined LOS models presented in Table 2 is misleading. None of the models is statistically significant at the .05 level or better (based on the Hosmer and Lemeshow Goodness-of-Fit test), and the proportion of variance explained is low (17% to 22%). Using either the DOC or the combined dataset, the models were not able to predict lengths of stay of less than 51 or 61 days. Moreover, most of the models’ seemingly impressive predictive accuracy can be attributed to the lopsided LOS distribution among prison-bound inmates. The vast majority of all prison-bound inmates stay at

Rikers for at least 51 days. In addition, as will be discussed below, apparent improvements in predictive accuracy that might otherwise be attributed to information from DOC or other sources must be considered suspect if only because of the conceptually erratic nature of the results. In any case, the predictive accuracy that one achieves with the help of the combined data set affords no improvement over the predictive accuracy that one achieves with DOC-only data.

Released to the Community Group

When it comes to predicting length of stay among inmates released to the community, the critical comparison is not between inmates who stay at Rikers for at least 37/45 days and those who stay at least 51/61 days. These two groups are virtually identical to each other. The critical comparison is between inmates who belong to either of these groups and the larger population of released inmates, which includes many inmates who leave Rikers in approximately one month or less. Thus, assessing the adequacy of the shorter LOS model, which encompasses all released inmates who stay at least 37 (or 45 days, depending on the data set) and compares them to the inmates with the shortest terms of less than 36 (or 45) days, is more important than assessing the adequacy of the longer LOS model—predicting stays of at least 51 (or 61) days as a separate entity.

Although none of the LOS models for released inmates explains a large proportion of the variance, as indicated in Table 2 by the values of R^2, all are statistically significant at the .02 level or better. Perhaps more important is the high level of accuracy that one achieves in predicting the shorter LOS with or without the additional detail provided by DCJS and CJA. As Table 2 shows, the DOC-only model correctly predicts a LOS of at least 37 days 81% of the time, while the combined data make the correct prediction of at least 45 days 77% of the time.

IV. Impact of Inmate Characteristics, Current Case Characteristics, and Criminal History on Outcome and Length of Stay

To make it easier to see patterns in the data, Table 3 summarizes the results of all ten logistic regression analyses, identifying the source of data for each predictor. The Table organizes the predictors into three categories, each shown on a separate page: inmate characteristics, characteristics of the current case, and criminal history. Where applicable, the cells identify the predictor category (such as gender or self-reported drug use) that is associated with each outcome (prison or release to the community) and LOS. The figures shown in the cells are odds ratios (OR) that are statistically significant; figures for nonsignificant predictors were omitted for readability. ORs shown with variables included in the DOC IIS were produced using analyses that included DOC data only. ORs shown with predictors from DCJS or CJA were produced using logistic regression analyses that employed the combined data set. Where applicable, the text will identify changes in the effects of DOC variables that occurred when the data from DCJS and CJA were included in the equations.

The discussion that follows concentrates on identifying which characteristics from each of the three data sources proved to be most important in predicting outcome and length of stay.

Predictors of Outcome: Prison or Release to the Community

Whether one relies on DOC data exclusively or on the combined data set, prison-bound and released inmates appear to have distinctive profiles. What most typifies the prison-bound is the relative seriousness of the charges currently against them and the presence of a prior record, often involving serious charges or disciplinary problems within the DOC system. Social background characteristics have only a small role in the profile of prison-bound inmates. The profile of inmates released to the community is very different. The current charges against them tend to fall on the low end of the severity spectrum. Although released inmates are not necessarily first-time offenders, any prior charges tend to be comparatively minor. Moreover, upon their release, these inmates can return, presumably, to an existing network of family and community ties their prison-bound counterparts tend not to have.

The relationship between borough of arraignment and outcome stand on their own conceptually, although what underlies these relationships is unclear. Specifically, having been arraigned in Manhattan predicts the prison outcome; having been arraigned in one of the other boroughs—especially Brooklyn, Queens, or the Bronx (OR between 1.27 and 2.18)—predicts release to the community.

Table 3: Predictors of Outcome and Length of Stay at Rikers: Summary of Odds Ratios from Logistic Regression Analyses

	Predictors	Source of Data	Outcome		Minimum Number of Days at Rikers for Inmates who are . . .			
					Prison-Bound		Released to the Community	
			Prison	Released to Community	Shorter LOS *(37/45 or >)*	Longer LOS *(51/61 or >)*	Shorter LOS (37/45or >)	Longer LOS *(51/61 or >)*
INMATE CHARACTERISTICS	**Sex**	**DOC**	Men ***(1.39)***	Women *(contrast.cat.)*				
	Citizenship	**DOC**				Non-U.S. ***(0.59)***	U.S. ***(1.43)***	U.S. ***(1.44)***
	Marital Status	**DOC**		Divorced or married ***(1.18, 1.33)***			Never married ***(contrast cat.)***	Never married ***(contrast cat.)***
	Living Arrangements	**DOC**						
	Drug Use (self-reported)	**DOC**	Drug use ***(1.29)***	No drug use ***(0.77)***			Drug use ***(1.22)***	Drug use ***(1.21)***
	Race/Ethnicity	**DOC**	Black or Hispanic ***(1.33, 1.40)***					
	Employment Status	**CJA**		Verified employment ***(1.21)***			No verified employment ***(contrast cat.)***	No verified employment ***(contrast cat.)***
	# Months at Current Address	**CJA**						
	CJA Stamp (release recomm. Based on community ties)	**CJA**		Release recom-mended ***(1.20)***		Qualified or not recomm. ***(1.27, 1.46)***		

. . . Continued on Next Page . . .

NOTES: Figures in cells are odds ratios (OR) that are statistically significant; figures for non-significant predictors were omitted for readability. Odds ratios shown with data from the NYC Department of Corrections (DOC) were produced using logistic regression analyses that included DOC data _only_. Odds ratios shown with data from the Division of Criminal Justice Services (DCJS) or the Criminal Justice Agency (CJA) were produced using logistic regression analyses that included data from _all_ sources.

Table 3, continued: Predictors of Outcome and Length of Stay at Rikers: Summary of Odds Ratios from Logistic Regression Analyses

	Predictors	Source of Data	Outcome		Minimum Number of Days at Rikers for Inmates who are . . .			
					Prison-Bound		Released to the Community	
			Prison	Released to Community	Shorter LOS (37/45 or >)	Longer LOS (51/61 or >)	Shorter LOS *(37/45 or >)*	Longer LOS *(51/61 or >)*
CURRENT CASE CHARACTERISTICS	**Borough of Arraignment**	**DOC**	Manhattan ***(contrast cat.)***	Bklyn, Qns, Bx, Other ***(1.27 – 2.18)***	Manhattan ***(contrast cat.)***	Queens ***(1.32),*** Manhattan ***(contrast cat.)***	Manhattan ***(contrast cat.)***	Manhattan *(contrast cat.)*
	Age at Admission	**DOC**			Older ***(1.02)***	Older ***(1.02)***		
	Bail Amount	**DOC**	> $1,000 ***(1.45 - >300.00)***	< $1,000 ***(.00 - .69)***	Relationship unclear	Relationship unclear	Higher bail ***(1.36 – 4.76)***	Higher bail ***(1.27 – 4.24)***
	Reman., Top Chg.	**DOC**		Yes ***(6.07)***				
	Top Severity Code	**DOC**		Low severity ***(1.15)***	High severity ***(0.70)***	High Severity ***(0.76)***	High severity ***(0.88)***	High severity ***(0.87)***
	Top Severity Category	**DOC**	Felony ***(1.83)***	Non-felony ***(0.55)***				
	Violent Top Charge	**DOC**				No violent off. ***(0.52)***	No violent off. ***(0.90)***	
	Property Offense as Top Charge	**DOC**	Property off. ***(1.64)***	No property off. ***(0.61)***		No Property Off. ***(.46)***		Property off. ***(1.34)***
	Person/Property Top Charge	**DOC**	Pers./Pro-Perty Chg. ***(1.71)***	No Pers./Pro-perty Chg. ***(0.59)***				
	Warrant Issued	**DOC**	Warrant ***(3.80)***	No warrant ***(0.31)***	Warrant ***(1.55)***		Warrant ***(3.12)***	Warrant ***(3.24)***
	Drug Offense as Top Charge	**DOC**	Drug Off. ***(1.66)***	No drug off. ***(0.60)***	No drug off. ***(0.24)***	No drug off. ***(0.16)***	No drug off. ***(0.83)***	
	Probation Eligibility	**DCJS**		Eligible ***(5.30)***	Not eligible ***(0.60)***	Not eligible ***(0.58)***		
	Child Victim	**DCJS**	Child vic. ***(1.27)***	No child vic. ***(0.80)***				
	DWI Arrest	**DCJS**	DWI ***(2.57)***	No (***0.39)***				
	Firearm/Weapon/ VFO Arrest	**DCJS**	F/W/V ***(1.26)***	No F/W/V ***(0.80)***	F/W/V ***(3.66)***	F/W/V ***(3.21)***		

Table 3, continued: Predictors of Outcome and Length of Stay at Rikers: Summary of Odds Ratios from Logistic Regression Analyses

	Predictors	Source of Data	Outcome		Minimum Number of Days at Rikers for Inmates who are . . .			
					Prison-Bound		Released to the Community	
			Prison	Released to Community	Shorter LOS *(37/45 or >)*	Longer LOS *(51/61 or >)*	Shorter LOS *(37/45 or >)*	Longer LOS *(51/61 or >)*
CRIMINAL HISTORY	**DOC Class. Cat.**	**DOC**	Low-med. to high (6–17+) ***(1.34 – 3.80)***	Low (0 – 5)			Low-med. to high (1.11 – 11.15)	Low-med. to high ***(1.14 – 17.67)***
	Prior Con. Category	**DOC**	Felony ***(1.96)***	Misdem., viol. ***(1.32)***		No Felony ***(0.79)***	Midem./viol. or felony ***(1.33, 1.26)***	Misdem./viol. or felony ***(1.19, 1.12)***
	Disciplinary History	**DOC**	Discipl. History ***(1.44)***	No Discipl. History ***(.69)***	Discipl. hist. (2.75)	Disciplinary history ***(2.91)***	Disciplinary history ***(1.43)***	Disciplinary history ***(1.41)***
	History of Violence	**DOC**		Violence ***(1.65)***				
	Prior Jail Term	**DCJS**						
	Prior Prison Term	**DCJS**	Prior prison term ***(1.34)***	No prior prison ***(0.75)***				
	Bench Warrant	**DCJS**						
	Felony Arrest	**DCJS**						
	Misdemeanor Arrest	**DCJS**						
	Drug Arrest	**DCJS**						
	Violent Felony Arrest	**DCJS**						
	Felony Convictions	**DCJS**						
	Misdemeanor Convictions	**DCJS**		Prior misd. ***1.01)***				
	Violation Convictions	**DCJS**				< convic.***(0.90)***		
	Infraction Convictions	**DCJS**				> convic. (***2.96)***		
	Drug Convictions	**DCJS**						
	Violent Felony Conv.	**DCJS**						
	Jail Sentences	**DCJS**						
	Prison Sentences	**DCJS**						
	Probation Sentences	**DCJS**		Probtn. ***(1.12)***				
	Time To Reoffense	**DCJS**				No reoff. or reoff. w/in past yr. ***(1.53)***	Reoffense w/in past yr. ***(contrast cat.)***	Reoffense w/in past yr. ***(contrast cat.)***

Prison-Bound Group

In the logistic regression analysis based on DOC-only data, prison-bound inmates tend to be male (OR=1.26) and report that they use drugs (OR=1.29), as Table 3 shows. Prison-bound inmates are also more likely to be black or Hispanic than white or members of another racial or ethnic group (OR=1.33, 1.40), but this finding should be considered more descriptive than predictive. In the later analysis based on the combined data set, race and ethnicity washed out as a predictor, suggesting that this variable is a cover for other factors statistically associated with race and ethnicity.

A current felony charge is predictive of a prison outcome (OR=1.93), as are charges for specific types of crime. These include offenses involving property damage (OR=1.64), property damage and person-related (OR=1.71) and drug s (OR=1.66). Judging from the comparatively high bail amounts and the presence of a warrant on the current charge (OR=3.80) that strongly predict a prison outcome, high risk of flight is another characteristic typical of inmates headed for state prison. The DOC data on inmates' criminal history underline the importance of the seriousness of past crimes, broadly speaking, as a predictor of prison. Higher scores on the DOC classification, particularly at the highest end, predict a prison outcome (OR between 1.34 and 3.80), as does a prior felony conviction (1.96) and a disciplinary history with DOC (1.44).[10]

The analysis based on the combined data set did not fundamentally alter the profile of prison-bound inmates. This analysis added to the list of current charges and types of crime that tend to predict a prison outcome, including crimes that involve a child victim (OR=1.27) and arrests for a DWI (2.57), for violent felony offenses, or offenses those involving firearms or weapons.(OR=1.26). The combined analysis also identified a prior prison term as a predictor of a prison outcome in the current case (OR=1.34).

Released to the Community Group

Inmates' social background characteristics play a larger role in predicting release to the community than to prison. In the analysis based on DOC-only data, being a woman is predictive of release, as is a personal history that includes marriage, regardless of whether the inmate is currently married or divorced (OR=1.18, 1.33). Inmates who are released tend to report <u>not</u> using drugs (OR=0.77).

Having been remanded on the top charge strongly predicts release to the community (OR=6.07). This finding might reflect the fact that charges against inmates who are released tend to be of low severity (OR=1.15) and do not involve a felony (OR=0.55). Released inmates tend to receive bail amounts under $1,000 (OR between .00 and 0.69) and did not have a warrant issued on the current charges (OR=0.31).

Consistent with the nature of the current case against them, a released inmate's criminal history, when one exists, typically consists of comparatively minor prior convictions, such as a misdemeanor or violation. A low score of zero to five on the DOC classification also predicts

[10] In the analysis that employed the combined data set, disciplinary history no longer predicted outcome.

release to the community, as does the absence of a disciplinary history within the DOC system (0.69). In a seeming anomaly, however, a history of violence predicts release (OR=1.65).

Although we cannot know the reason for this striking departure from the broader pattern in the data, this finding might reflect a disproportionate number of domestic violence cases among released inmates. (If so, however, these cases most likely do not include a disproportionate number of instances of violence against a child; see below.)

The data from DCJS and CJA add some detail to the portrayal offered by DOC data alone. CJA information indicating verified employment and a CJA release recommendation both predict release to the community (OR=1.21, 1.20), reinforcing the impression of relatively conventional personal histories that emerges from the DOC data. DCJS data underline the less serious nature of the offenses committed by inmates who are released to the community and their less eventful criminal histories. Eligibility for parole, for example, is among the strongest predictors of release (OR=5.30). Also predictive of release, according to DCJS data, are the absence of a child victim (OR=0.80) and, the absence of an arrest for DWI offenses (OR=0.39) or for violent felony offenses or offenses related to firearms or weapons(OR=0.80). Consistent with the broader picture, DCJS data predict release for inmates who have not served a prior prison term (OR=0.75), whose prior convictions (if any) involve a misdemeanor, and who have previously been sentenced to probation (OR=1.12).

V. Predictors of Length of Stay

Prison Bound Group

Neither the DOC data nor the combined data set reveal clear and consistent differences between prison-bound inmates with the longest stay of at least 51 or 61 days and those with the shorter stay of at least 37 or 45 days, as Table 3 shows. Nor, for that matter, are there clear and consistent differences between these two prison-bound groups and the larger sample of Rikers inmates who went to prison (which includes the few prison-bound inmates with the shortest stays of 36 days or less). The absence of a discernable pattern in the data underscores the problems inherent in attempting to identify reliable predictors of stays of at least 37 or 45 days on the one hand and stays of at least 51 or 61 days on the other for prison-bound inmates.

In some respects, both of these groups are very similar to the larger sample of prison-bound inmates (with respect to borough of arraignment, seriousness of offense, and having a disciplinary history with the DOC system). The addition of DCJS data reinforces this pattern to the extent that they identified ineligibility for parole and an arrest for a violent felony offense or for offenses involving a firearm or weapon as predictors of membership in both the shorter and the longer LOS groups.

Other findings appear to point in the other direction, suggesting that a longer LOS for prison-bound inmates is associated with less severe offenses, or point to opposite conclusions. For instance, DOC data indicate that the absence of a prior felony conviction predicts longer LOS (OR=0.79), as does a nonviolent top charge (OR=0.52) and a top charge that does not involve a property offense (OR=0.46). Similarly, judging from DCJS data, having a low number of prior violation convictions predicts a longer LOS for prison-bound inmates (OR=2.96). A longer LOS for prison-bound inmates, however, is associated with increasing numbers of convictions for infractions.

Despite this inconsistent picture, several findings might shed some light on the social background characteristics of prison-bound inmates whose stay at Rikers was among the longest. DOC data indicate that being a citizen of a country other than the United States predicts a longer LOS (OR=.59) and that prison-bound inmates in both LOS groups (whose stay at Rikers lasts more than 37 days) tend to be a bit older than the few inmates who leave Rikers very quickly (OR=1.02). In addition, a longer LOS is associated with tenuous community ties in that CJA provided only a qualified recommendation for release or recommended against release (OR=1.27, 1.46).

Released to the Community Group

In contrast to the seemingly erratic findings about LOS among prison-bound inmates, comparable findings for released inmates reveal two striking patterns. First, in almost all respects, the predictors of shorter and longer lengths of stay are identical. Second, released inmates with shorter <u>and</u> longer lengths of stay tend to have a different profile than the larger sample of inmates who were released to the community. In some respects, inmates in both LOS groups (all of whom stayed at Rikers for at least 37 or 45 days, depending on the data source) have less in common with the inmates who returned to the community more quickly than with the larger population of inmates who are headed for prison. On the basis of the available data, then, one cannot predict which released inmates are most likely to stay at Rikers for at least 37 or 45 days and which are the most likely to stay for at least 51 or 61 days. What the available data <u>can</u> do is identify factors that differentiate those likely to stay at Rikers for at least 37 or 45 days from the broader population of inmates who are released to the community, often very quickly. Most of the data that contribute to making this differentiation come from DOC.

Released inmates in both the shorter and longer LOS groups, like prison-bound inmates, tend to have committed fairly serious crimes. The top severity code for both LOS groups of released inmates is high (OR=0.88, 0.87), as is the bail amount (OR between 1.36 and 4.76; and between 1.27 and 4.24). Both LOS groups have a low-medium to high DOC classification (OR between 1.11 and 11.15; and between 1.14 and 17.67) and a disciplinary history within the DOC system (OR=1.43, 1.41). Also like the prison-bound population, having committed a prior felony (OR=1.26, 1.12, as well as a prior misdemeanor or violation, OR=1.33, 1.19) predicts an LOS of at least 37 (or 45) days, as does having a warrant issued in the current case. Further, arraignment in Manhattan predicts membership in one of the LOS groups of released inmates just as it predicts a prison outcome.

Unlike the larger sample of inmates released to the community, which includes those released very quickly, both LOS groups of released inmates have social background characteristics indicative of tenuous community ties. Having never been married predicts membership in these LOS groups whereas, as we have seen, a marital history is typical of the broader population of Rikers inmates who are released to the community. And, like prison-bound inmates in general, members of both LOS groups of released inmates report drug use (OR=1.22, 1.21), although (unlike the situation for many inmates with a prison outcome), drug offenses did not necessarily play a part in determining the charges against them; having a top charge that did not involve a drug offense predicts shorter LOS for inmates released to the community (OR=0.83).

Once again, data from DCJS and CJA augment the portrayal offered by DOC data without fundamentally changing it. For example, consistent with other indicators of tenuous community ties, released inmates who belong to both LOS groups have no verified employment according to CJA. Having committed a reoffense within the past year predicts membership in both LOS groups of released inmates, according to DCJS data. Like the presence of a warrant on the current case, this finding suggests that a determination of flight risk helps to distinguish between released inmates who belong to either of the LOS groups and inmates who are released to the community more quickly.

Three findings appear to stand independently in that they do not fit into the patterns just described. U.S. citizenship predicts membership in both LOS groups of released inmates (OR=1.58, 1.56) but is not a factor in predicting outcome. The absence of an offense involving violence predicts shorter LOS for released inmates but, like citizenship, does not play a part in outcome. Finally, the absence of a top charge involving a drug offense (OR=0.83) is one respect in which released inmates with a shorter LOS are similar to released inmates in general and is one factor (along with the absence of a top charge involving violence) that distinguishes released inmates in the shorter LOS group from released inmates in the longer LOS group.

Chapter Three

Survival Analysis of Length of Stay at Rikers

Logistic regression analysis (LRA) is a useful and reasonably accurate method of predicting the outcome of prison versus the outcome of release to the community. But it proved to be less useful in predicting shorter lengths of stay for the prison-bound group and longer lengths of stay for the released to the community group. For predicting length of stay (LOS), LRA is inflexible because it forces one to construct a single fixed cut-off point that divides the inmate population into just two categories; inmates stay at Rikers for a specified minimum length of time or they do not. During the planning phase of this research, survival techniques appeared to offer a more promising approach to predicting LOS. The predictions generated by survival analysis, however, probably do not constitute an improvement over the predictions generated by LRA.

I. Survival Analysis

Survival analysis examines "time until failure," to use the statistical terminology, and the nature of that failure. In the present analysis, "time until failure" means number of days incarcerated in Rikers or, stated differently, number of days until inmates leave Rikers for state prison or the community. Survival analysis offers more flexibility than LRA in that it estimates the number of days, whatever that might be, instead of limiting the investigators to specific LOS categories. One can, but need not, assess the model's usefulness based on its ability to predict a length of stay that is longer or shorter than a specified cut-off point.

For several reasons, the survival analysis results reported below should be considered preliminary. Unlike the logistic regression analyses, the survival analyses rely solely on data provided by DOC. Since the use of the combined dataset did not markedly improve the predictive accuracy of the logistic regression models, we decided to conduct the survival analysis using DOC data only. Applying survival analysis to a combined data set, of course, remains an option for future research. Another option is to look for interaction effects among predictors, which we did not do here.

We first discuss which characteristics the survival analyses identified as predictors of LOS among prison-bound inmates and released inmates. This chapter concludes with an assessment of the adequacy of the models.

II. Predicting LOS For Prison-Bound Group

Using the exponential distributional form, we employed survival analysis to predict length of stay for prison-bound inmates. The analysis is based on the 7,440 cases in a slightly modified version of the DOC-only data set.[11] An exponential regression model, commonly used to predict

[11] One change to the data was the creation of a dummy variable indicating whether or not information was missing on charge severity (coded 1 if information was missing and 0 if it was not). This addition made it possible to retain

the time to an event, here predicts length of stay. The model assumes that the probability of release is constant over time. Although likelihood ratio tests and inspection of the AIC suggest that a lognormal model might have fit the data better, we chose the exponential model because it was the only one to identify any prison-bound inmates with a length of stay less than 45 days. Prior to estimating the LOS, the clock was reset by subtracting ten from the length of stay, producing a model that predicts the length of time until release from the eleventh day onward.

Table 4 summarizes the resulting prediction equation. The column labeled “B” shows the maximum likelihood estimated coefficients. These coefficients are analogous to slopes (unstandardized coefficients) estimated in ordinary least squares regression. They are the estimated contribution to the LOS associated with a given variable when other variables in the equation are held constant. Table 4 also shows the estimated standard error (i.e., the variability around the estimated coefficient) and the significance level (*p*) for each parameter estimate. Values below 0.05 indicate that a relationship is statistically significant at the 95% confidence level; values below 0.10 indicate that a relationship is statistically significant at the 90% confidence level. The fifth and final column reports the percentage increase or decrease in the estimated length of stay attributable to each variable in the model.

---

the 211 cases with missing information on this variable. Another modification created a dummy variable for top charge that indicated a charge other than one involving violence, drugs, or property. This variable was coded 1 if the top charge was something other than these three, and 0 of it one of the three. Finally, age at admission was recoded to truncate the values to the integer year.

Table 4: Prediction Equation for LOS Among Prison-Bound Group

Variable	B	Standard Error	*p*	% Change
Manhattan	-0.208	0.036	0.000	-18.8
Brooklyn	-0.247	0.043	0.000	-21.9
Bronx	-0.229	0.040	0.000	-20.5
Staten Island	-0.454	0.104	0.000	-36.5
Other Arrangements	-0.205	0.068	0.002	-18.6
Male	0.129	0.042	0.002	13.8
US Citizen	-0.062	0.039	0.106	-6.0
African-American/Non-Hispanic	0.050	0.025	0.041	5.2
Felony B	-0.565	0.041	0.000	-43.2
Felony C	-0.779	0.055	0.000	-54.1
Felony D	-0.863	0.060	0.000	-57.8
Felony E	-0.773	0.083	0.000	-53.8
Misdemeanor A	-0.735	0.081	0.000	-52.1
Misdemeanor B	-1.038	0.126	0.000	-64.6
Missing Top Severity	-1.186	0.109	0.000	-69.5
Bail 5,001+	0.055	0.030	0.071	5.6
Warrant Present	-0.122	0.043	0.004	-11.5
Violent	-0.107	0.070	0.123	-10.2
Drug	-0.737	0.070	0.000	-52.1
Property	-0.393	0.074	0.000	-32.5
DOC Classification Score	0.038	0.008	0.000	3.9
Classification Score Greater than 15 (yes/no)	0.515	0.086	0.000	67.3
Misdemeanor Prior	-0.185	0.042	0.000	-16.9
Felony Prior	-0.210	0.047	0.000	-18.9
Remanded	-0.278	0.032	0.000	-24.3
Disciplinary History	0.290	0.058	0.000	33.7
Violence History	-0.180	0.045	0.000	-16.4
Age (years)	0.013	0.002	0.000	1.3
Constant	5.626	0.127	0.000	NA

The survival analysis of LOS among prison-bound inmates produced a somewhat more coherent profile of this group than the logistic regression analyses (LRA), even without the benefit of the additional data from DCJS and CJA. Nevertheless, many findings in both analyses are either very similar or conceptually consistent with each other.

With respect to inmates' background characteristics, for instance, the survival analysis more strongly suggests a connection between minority status, generally speaking, and longer LOS. Like the logistic regression analysis, for example, the survival analysis ties non-U.S. citizenship to longer lengths of stay; as Table 4 shows, the stay for U.S. citizens was 6% shorter than for citizens of other countries. Table 4 also shows that the LOS for prison-bound African Americans was 5.2% longer than for other racial and ethnic groups included in the analysis. The survival analysis also provided more detail about inmates' social background characteristics to the degree that it ties longer LOS to being male (percent change = 13.8%), whereas inmates' gender did not emerge as a predictor of LOS in the LRA results. Finally, like the LRA, the survival analysis links older age at admission to Rikers to longer LOS; for every increase of one year in inmates' age, there is a 1.3% increase in the predicted length of confinement. Both statistical techniques associate arraignment in Queens with longer LOS.

As in the logistic regression analysis of LOS among prison-bound inmates, the survival analysis yields somewhat inconsistent results about the impact of current case characteristics and criminal history. Inmates with a current top charge of Felony A tend to have considerably longer lengths of stay than do inmates with lower-severity top charges. Being charged with a Felony B crime rather than a Felony A, for instance, reduces length of stay by 43.2%. Consistent with a connection between more serious crimes or flight risk and longer LOS, cases involving a bail amount of at least $5,001 remained at Rikers 5.6% longer than cases with lower bail amounts. On the other hand, time served at Rikers was 11.5 % shorter for cases in which the top charge does not involve a violent crime (10.2% shorter), drugs (52.1% shorter), or a property offense (32.5% shorter). Consistent with findings suggesting shorter stays for less serious current offenses, the LOS is 16.9% shorter among inmates previously convicted of a misdemeanor than among inmates with no prior conviction, 18.9% shorter for inmates previously convicted of a felony than for people never previously convicted of a crime, and 16.4% shorter for inmates who have a history of violence than for those who do not.

Although the nature of inmates' crimes has inconsistent effects on LOS, previous problems within the DOC system strongly predict longer stays. A preliminary analysis showed that higher DOC classification scores predict longer confinements, but the relationship associated with scores higher than 15 is even larger than the linear form of the score would suggest. Therefore, two variables—DOC score and dichotomous variable that indicates the presence or absence of a DOC score is higher than 15—were included in the model. The results suggest that for every increase of one point in the DOC classification score, the LOS increases by 3.9%. When the score exceeds 15, however, LOS increases by an additional 67.3%. Similarly, LOS is 33.7% longer for inmates who have a disciplinary history than for inmates who do not.

III. Predicting LOS Among Released to the Community Group

The survival analysis of LOS among released inmates is based on a slightly modified version of the DOC data set and includes 16,435 cases.[12] We predicted LOS using the Weibull regression

---

[12] As in the analysis for the prison-bound, a dummy variable was created to indicate whether or not information was missing on charge severity, making it possible to retain the 382 cases with missing information on this variable.

model, which assumes that the probability of release is highest initially and decreases monotonically over time. Prior to estimating time, the clock was reset by subtracting ten days from the lengths of stay, so that it predicts length of time until release from the eleventh day onward. Table 5 shows the resulting equation, which employed the shape parameter to determine the particularly Weibull function fit to the data. The estimated value is 0.870 which, like all values less than 1.0, indicates that the probability of release decreases over time.

Another modification created a dummy variable for top charge that indicated a charge other than one involving violence, drugs, or property. Finally, age at admission was recoded to truncate the values to the integer year.

Table 5: Prediction Equation for LOS Among Released to the Community Group

Variable	B	Standard Error	*p*	% Change
Manhattan	0.322	0.024	0.00	37.9
Bronx	0.244	0.028	0.00	27.7
Queens	0.168	0.028	0.00	18.3
Staten Island	0.292	0.072	0.00	33.9
Other Arrangements	0.195	0.049	0.00	21.5
Male	0.125	0.029	0.00	13.3
U.S. Citizen	0.199	0.033	0.00	22.0
Single, Never Married	0.086	0.023	0.00	9.0
Drug Use	0.108	0.025	0.00	11.4
African-American/Non-Hispanic	0.038	0.019	0.05	3.8
Felony A	0.272	0.072	0.00	31.3
Felony B	0.358	0.043	0.00	43.1
Felony C	0.338	0.042	0.00	40.2
Felony D	0.222	0.036	0.00	24.8
Felony E	0.226	0.042	0.00	25.3
Missing Top Severity	0.360	0.068	0.00	43.3
Bail $501-1,000	0.108	0.035	0.00	11.4
Bail $1001-2,500	0.309	0.032	0.00	36.1
Bail $2,500-5,000	0.534	0.036	0.00	70.6
Bail $5,001-10,000	0.418	0.038	0.00	51.9
Bail $10,001-50,000	0.729	0.048	0.00	107.4
Bail $50,001+	1.058	0.085	0.00	188.0
Warrant Present	0.408	0.042	0.00	50.4
Violent	-0.058	0.034	0.09	-5.6
Drug	-0.162	0.038	0.00	-15.0
Other	-0.103	0.034	0.00	-9.8
DOC Classification Score	0.072	0.006	0.00	7.5
Class. Score Greater than 15 (yes/no)	0.650	0.146	0.00	91.5
Misdemeanor Prior	0.052	0.028	0.06	5.3
Felony Prior	-0.165	0.028	0.00	-15.2
Remanded	0.069	0.025	0.01	7.2
Disciplinary History	0.171	0.049	0.00	18.7
Violence History	-0.268	0.037	0.00	-23.5
Age (years)	0.009	0.001	0.00	0.9
Constant	2.388	0.083	0.00	NA
shape parameter p	0.870	0.005		

In many respects, the findings from the survival analysis for inmates released to the community parallel the findings produced by the logistic regression analyses (LRA). For example, released inmates arraigned in Manhattan had the longest stays--37.9% longer than people arraigned in Brooklyn, who had the shortest stays. In the survival analysis, as in the LRA, U.S. citizens stay at Rikers longer (22% longer) than noncitizens. The survival analysis suggests, as did the LRA, that released inmates with longer lengths of stay seemed more similar, in many respects, to prison-bound inmates than they did to the released inmates with the briefest stays of 36 days or less. For example, as for prison-bound inmates in general, longer lengths of stay among released inmates are more characteristic of men than of women (9% longer), of African-Americans than of other racial and ethnic groups (3.8% longer), of single or never-married inmates than of inmates who live with others (9% longer), and of inmates who report drug use than of those who do not (11.4% longer). And, similar to prison-bound inmates in both the longer and shorter LOS groups in the logistic regression analysis, survival analysis indicates that for every increase of one year in age, there is a 0.9% increase in the expected length of confinement at Rikers.

As in the LRA, released inmates with longer lengths of stay also tend to be similar to prison-bound inmates in the seriousness of current charges. A longer LOS is associated with arraignment on charges more serious than a misdemeanor. People arraigned on a Felony A charge stayed at Rikers 31.3% longer. The LOS for released inmates arraigned on Felony B, C, and E charges were longer than for people charged with a misdemeanor by 43.1%, 40.2%, 24.8%, and 25.3% respectively. As in the LRA findings, LOS is longest for inmates in which the top charge involves a property crime; the LOS for released inmates with charges involving violence, drugs, and other types of crime is shorter than for property crimes by 5.6%, 15%, and 9.8% respectively. Consistent with these findings, and with the role of flight risk in shaping LOS, the presence of a warrant on the current charge increases LOS by 50.4%, and LOS increases dramatically with increases in the bail amount. For example, by comparison to releases whose bail was set at $500 or less, LOS increased 11.4% for bail amounts ranging from $501 to $1,000. When the bail exceeded $10,000, the LOS doubled (i.e., an increase of more than 100% a bail amount in excess of $50,000 increased released inmates' LOS by 188%, almost tripling the LOS by comparison to inmates with the lowest bail amounts.

The effects of criminal history on released inmates' LOS are somewhat mixed. The LOS for released inmates previously convicted of a felony is 15.2% shorter than for inmates never previously convicted of a crime, and LOS is 23.5% shorter than it is for inmates with no history of violence. The LOS for inmates previously convicted of a misdemeanor, however, is 5.3% longer than for inmates with no prior convictions.

The survival analysis of LOS among released inmates, like other findings in this study, points to the importance of previous experiences with DOC in shaping length of stay. For every increase of one point in the DOC classification score, the LOS for released inmates increases by 7.5%. When the score exceeds 15, the LOS nearly doubles, increasing by an additional 91.5%. Similarly, the number of days inmates are detained at Rikers is 18.7% longer for inmates with a disciplinary history than for inmates without such a history.

IV. Assessing the Adequacy of the Survival Analyses for Predicting Length of Stay

To gauge the adequacy of the survival analysis models for predicting LOS among prison-bound inmates and among inmates released to the community, we looked at percentage of variance explained and the percentage of cases correctly predicted. Because of DOC's interest in predicting which inmates are likely to remain at Rikers long enough to complete a SAID program of 45 days, we examined the survival analysis' success in predicting lengths of stay at a 55 day cutoff, or 45 days after the SAID assessment on day 10. For comparative purposes, we also assessed the models according to their ability to predict stays at a 45 day cutoff, or 35 days after the SAID assessment). Because the survival model predicts LOS in actual number of days, however, any cutoff point can be assessed. One need only recode the actual and predicted times to form dichotomous indicator variables reflecting the desired cutoff point, adjusting the predicted times by adding ten days prior to dichotomizing. Table 6 summarizes the findings for the cutoff points we selected.

Table 6: Gauges of Adequacy for the Survival Analysis Models

Gauges of Adequacy	**Length of Stay for Inmates who are . . .**	
	Prison-Bound	**Released to the Community**
% of variance explained	30.8% (p <.001)	16.9% (p <.001)
% cases correctly predicted when LOS is:		
Less than 45 days	3.4%	64.8%
45 days or more	89.2%	60.0%
Total	**88.9%**	**62.4%**
Less than 55 days	8.58%	83.6%
55 days or more	97.8%	34.6%
Total	**81.2%**	**61.0%**

Prison-Bound Group

In two respects, the survival analysis of LOS among prison-bound inmates appears to constitute an improvement over the logistic regression analysis. First, the survival analysis arguably provides a somewhat conceptually clearer picture of prison-bound inmates with longer versus shorter lengths of stay. Second, as Table 6 shows, the survival analysis explains more of the variance (30.8%, compared to 17% as shown in Table 2 in the previous chapter). Although the survival analysis does not do quite as good a job as the logistic regression analysis did at correctly predicting longer stays, it nevertheless made the correct prediction most of the time, for 89.2% of prison-bound inmates whose stay at Rikers lasted for at least 45 days and for 97.8% of the inmates whose stay lasted at least 55 days.

Nevertheless, using survival analysis to predict LOS for prison-bound inmates yields little return for the effort. That analysis, like the logistic regression analysis, only slightly improves on predicting a minimum stay of 45 or 55 days for all prison-bound inmates, 92% of whom stay at Rikers for at least that long (refer to Table 1). Note too that our survival analyses did an extremely poor job of predicting shorter lengths of stay for prison-bound inmates, predicting correctly between 3.4% and 8.5% of the cases depending on which cutoff point was used.

Released to the Community Group

With respect to the percentage of the variance explained, the survival analysis performed slightly better than the logistic regression analyses (LRA) for released inmates, but both performed moderately well. As Table 6 shows, the survival analysis explained almost 17% ($p<.001$) of the variation in LOS for inmates released to the community, reflecting a correlation between the predicted and actual time of 0.412; the LRAs, as Table 2 demonstrated, explained between 11% and 15% of the variance using DOC-only data.[13]

When using the 45-day cut off, the survival analysis correctly classifies almost two-thirds of the released inmates who stay at Rikers for less than 45 days and for 45 days or longer (64.8% and 60.0% respectively). When using the 55-day cut off, the survival analysis correctly predicts almost 84% of cases involving released inmates who leave Rikers before 55 days have elapsed. In stark contrast, it correctly classifies only one-third of released inmates whose stay at Rikers lasts for 55 days or more. Although the specific figures are different, these findings are consistent with the LRA results, which also showed more accurate predictions for shorter-term than for longer-term released inmates (as shown in Table 2). Not shown in Table 6 is that the model performed especially poorly in classifying very long stays of more than 100 days. In fact, the model predicts an LOS in excess of 65 days only 10% of the time.

[13] These comparisons are only approximate because of modest differences in the cut-off points used in the LRAs and the survival analysis.

Chapter Four

Conclusions

Using statistical predictions to improve the fit between the SAID program and the differing needs of prison-bound inmates and released inmates provides some constructive lessons. The models we employed are valuable tools in predicting outcome and length of stay, although not always to the extent, or in the precise ways, that we had anticipated.

We were able to identify specific predictors of prison or release to the community outcomes, one of the central tasks that we set out to accomplish. Predicting LOS, however, proved to be more difficult, and it requires a more nuanced understanding of the findings than is necessary for predicting outcome. In particular:

- The LOS categories we employed (because of their particular interest to DOC administrators) make programmatic sense in that they reflect realistic durations of treatment. They do not, however, lend themselves to making reliable LOS predictions. Although the broad subgroups of prison-bound inmates and released inmates tend to be very different from each other (making it possible to predict outcome), <u>inmates with shorter and longer lengths of stay as defined in this report do not have distinctive profiles</u>.

- <u>LOS for prison-bound inmates does not vary very much</u>.The vast majority of these inmates have, for purposes of SAID programming, lengths of stay of 51 days or more.

- Almost half of all released inmates leave Rikers within approximately one month of admission. <u>Released inmates who stay at Rikers long enough to benefit from any treatment intervention</u>, whether short-term or intensive, <u>have more characteristics in common with prison-bound inmates than they do with the broader population of inmates who are released to the community</u>. For released inmates, then, it <u>is</u> possible to predict which of them are the most plausible candidates for at least a minimal intervention. On the basis of the available data and cutoff points, however, it is <u>not</u> possible to predict which released inmates will stay at Rikers long enough to complete intensive treatment.

Based on these conclusions and various concrete findings, we outline below a proposed series of steps for improving the fit between SAID interventions and inmates' particular needs.

Step 1: Selecting the Source of Data

Using the combined data set, which included information from DCJS and CJA as well as DOC, did not fundamentally alter the predictors or the predictive accuracy that we obtained from using data from DOC exclusively. If DOC were to invest the time and resources in using the combined data set, this decision would have the most, albeit limited, pay-off in predicting the prison outcome. This is the only logistic regression analysis that produced modestly better predictive results with the combined data than with the DOC- only data set.

Step 2: Predicting Outcome

The logistic regression analysis achieved a fair degree of accuracy in predicting the outcomes of prison and especially release to the community. Broadly speaking, however, we found that prison-bound inmates are characterized less by their social background characteristics than by (1) the seriousness of the current charges against them and (2) the presence of a prior record, which often involves serious charges or disciplinary problems within the DOC system. In contrast, inmates released to the community tend to have (1) an existing network of family and community ties, (2) been charged with offenses that fall on the low end of the severity spectrum, and (3) a prior criminal record, if any, that involves comparatively minor charges. (See Table 3 for more detail.)

Step 3: Assuming a Long LOS for Prison-Bound Inmates

Predicting a prison outcome is virtually tantamount to predicting a relatively long LOS of at least 51 days (using the DOC-only data set). Eighty-four percent of prison-bound inmates stayed for 51 days or more, and an additional 10.6 % stayed between 37 and 50 days. Only 5.3% left Rikers in 36 days or less. One can only marginally improve on this predictive accuracy by employing either the logistic regression or survival analyses of LOS. DOC administrators will have to decide whether the marginal increase in predictive accuracy that one can achieve through statistical procedures is worth the investment of time and resources that performing these procedures would entail. We suspect that it is not, and suggest that DOC staff use the statistical procedures to predict the prison-bound outcome and assess this group for an intensive SAID program.

Step 4: Predicting LOS among Inmates Released to the Community

In predicting LOS for released inmates, the critical comparison is between inmates who stay at Rikers for at least 37 days and those who leave in 36 days or less. Neither the logistic regression analysis nor the survival analysis succeeded in distinguishing between inmates with the “shorter” LOS of at least 37 days (using DOC-only data) and the “longer” LOS of at least 51 days. The survival analysis in particular made it abundantly clear that predictive accuracy for released inmates diminishes as the LOS that one is trying to predict increases.

As we have seen, the “shorter” and “longer” LOS released inmates (with minimum stays of 37 and 51 days respectively, using DOC data exclusively) are virtually identical to each other in social background characteristics, characteristics of the current case, and criminal history. Like prison-bound inmates as a group, released inmates who are likely to stay at Rikers at least 37 days tend to have rather tenuous family and community ties and to have committed serious offenses. Our analysis did not specifically describe the shortest LOS subgroup as a separate entity, but it suggests that they have rather conventional social backgrounds, in some respects, and have been detained for comparatively minor offenses.

In predicting LOS for released inmates, one can use either logistic regression analysis (LRA) or survival analysis and achieve similar predictive results. Decisions about which of these statistical techniques to employ depend on the importance that administrators attach to how straightforward it is to use these techniques without having to resort to sophisticated and expensive statistical software packages (such as SAS or SPSS), a consideration that favors the use of LRA, or the flexibility that survival analysis affords in experimenting with the utility of various cutoff points. The appendix describes steps in implementing predictions using each of these techniques, which should assist administrators in deciding which technique best meets their needs.

If one cannot reliably predict which released inmates are likely to stay at Rikers long enough to complete participation in an intensive intervention, where does this leave us? For released inmates likely to stay at Rikers for at least 37 days, is there no alternative to the current practice of offering a uniform SAID program to all inmates, knowing that many will return to the community before completing treatment? Below, we suggest an alternative to the status quo.

Step 5: Revisiting some Assumptions about the Structure of SAID Interventions for Released inmates

In a preliminary response to a draft of this report, DOC administrators envision two well-defined substance abuse treatment programs, both based on the model of the therapeutic community. The shorter-term programs, geared to inmates who will be released to the community within approximately 55 days after admission to Rikers, would concentrate on discharge plans supplemented by "group interventions that motivate the participants to continue treatment after discharge." The more intensive program, geared to inmates likely to be released to the community or to prison after a longer stay, would be more treatment oriented. Inmates would receive "individual and group counseling with progressive treatment outcomes and discharge planning leading to similar programs either in State prison or in the community."

Because of the impossibility of predicting which released inmates will stay at Rikers longer than 37 days, this dual program might not be feasible. The substance of the interventions need not necessarily change fundamentally, however, if treatment programs are specifically designed to take into account the unpredictability that surrounds released inmates' LOS at Rikers. For example, all treatment curricula could begin with as much discharge-oriented content as possible, leaving the more complex and time-consuming therapeutic tasks for later in the course of treatment. In general, by creating curricula that consist of discrete treatment modules, each of which can stand alone to the greatest extent possible, administrators can help to compensate for the problems inherent in trying to predict length of stay for inmates who ultimately return to their communities.

References

Belenko, S., Phillips, M.T., Demetriades R., Eliyahu D., Winterfield, L., & Heffernan, R. (1993). Predicting Incarceration Length: Estimating the Displacement Effects of Alternatives-To-Incarceration Programs: Final Report of the Jail Displacement Study. New York: New York City Criminal Justice Agency.

Belenko, S., Schiff, M.F., Phillips, M.T., & Winterfield, L.A. (1994). Modeling the Prison Displacement Effects of Alternatives Sanction Programs: A Case Study. *The Prison Journal*. 73(2): 167-197.

Belenko, S., Winterfield, L., Phillips, M.T., Grant, A.C., & Caligiure, R. (1995). Estimating the Displacement Effects of Alternatives-To-Incarceration Programs: Final Report of the Jail Displacement Study Phase II. New York: New York City Criminal Justice Agency.

Center on Addiction and Substance Abuse. (1998). Behind bars: substance abuse and America's prison population. New York: CASA.

Clear, T. (1988). Statistical Prediction in Corrections. *Research in Corrections.* Washington D.C.: The National Institute of Corrections.

Daly, Kathleen & Michael Tonry (1997) Gender, Race, and Sentencing. In M. Tonry (Ed.) *Crime and Justice: A Review of Research Volume 20,* Chicago: The University of Chicago Press.

Gaes, G., T. Flanagan, L. Motiuk, L. Stewart (1999) Adult Correctional Treatment. In Michael Tonry, Joan Petersilia (Eds.) *Prisons*, Chicago: The University of Chicago Press.

Glaser, D. (1987). Classifications for Risk. In D.M. Gottfredson & M. Tonry (Eds.) *Prediction and Classification, Criminal Justice* Decision *Making, Crime and Justice Volume 9*, Eds. Chicago: The University of Chicago Press.

Gottfredson D.M. (1987). Prediction and Classification in Criminal Justice Decision Making. In D.M. Gottfredson & M. Tonry (Eds.) *Prediction and Classification, Criminal Justice Decision Making, Crime and Justice Volume 9*, Chicago: The University of Chicago Press.

Hepburn, J.R. (1994) Classifying Drug Offenders for Treatment. In D.L. McKenzie & C. D. Uchida (Eds.) *Drugs and Crime: Evaluating Public Policy Initiatives*. California: SAGE Publications, Inc.

Field, G. (1989). The effects of intensive treatment on reducing the criminal recidivism of addicted offenders. *Federal Probation*, 51-56.

Harlow, C.W. (1998). Profile of jail inmates 1996. Bureau of Justice Statistics Special Report. Washington, D.C: Dept. of Justice.

Inciardi, J.A., Martin, S.S., Butzin, C.A., Hooper, R.M., & Harrison, L.D. (1997). An effective model of prison-based treatment for drug-involved offenders. *Journal of Drug Issues*, 27 (2), 261-278.

Knight, K., Simpson, D.D., & Hiller, M.L., Three-year reincarceration outcomes for in-prison therapeutic community treatment in Texas. *The Prison Journal*, 79, 337-351.

Lattimore, P.K., & Linster, R.L. (1996). Risk Classification of Probationers: Development of a Statistical Model Based on Management Information System Data. National Institute of Justice Working Paper 96-01.

Lipton, D.S. (1996). Prison-Based Therapeutic Communities: Their Success with Drug- Abusing Offenders. *National Institute of Justice Journal* 230:12-20.

Lipton, D. S. (1995). The effectiveness of treatment for drug abusers under criminal justice supervision. National Institute of Justice Report. Washington: Department of Justice.

Martin, S.S., Butzin, C.A., Saum, C.A., & Inciardi, J.A., Three-year outcomes of therapeutic community treatment for drug-involved offenders in Delaware: from prison to work release to aftercare. *The Prison Journal*, 79, 294-320.

Morris, N. & Miller, M. (1985). Predictions of Dangerousness. In M. Tonry & N. Morris (Eds.) *Crime and Justice: An Annual Review of Research*, Vol. 6. . Chicago, University of Chicago Press.

Mumola, C.J. (1999). Substance abuse and treatment, state and federal prisoners, 1997. Bureau of Justice Statistics Special Report. Washington: Dept. of Justice.

New York State Department of Correctional Services (1999) The Comprehensive Alcohol and Substance Abuse Treatment Program. Albany, NY: New York State Department of Correctional Services.

Peters, R., Kearns, W., Murrin, M., Dolente, Addis, & May, R. (1993). Examining the Effectiveness of In-Jail Substance Abuse Treatment. *Journal of Offender Rehabilitation* 19: 3/4.

Peters, R. H. (1993). Drug Treatment in Jails and Detention Settings. In J. Inciardi (Ed.) *Drug Treatment and Criminal Justice*. Sage Criminal Justice System Annuals, Vol 27. California: Sage Publications.

Peters, R.H., & Steinberg, M.L. (2000). Substance abuse treatment services in U.S. prisons. In D. Shewan & J. Davies (Eds.), *Drugs and Prisons*, 89-116. London: Harwood Academic Publishers.

Sims and Jones (1997) Predicting success or failure on probation: factors associated with felony probation outcomes. *Crime and Delinquency* 43 (3): 314-327.

Tunis, S., Austin, J., Morris, M., Hardyman, P., & Bolyard, M. (1996). Evaluation of Drug Treatment in Local Corrections. A Final Report Submitted to the National Institute of Justice. Washington: Dept. of Justice.

Walsh, "Teaching Understanding and Interpretation of Logit Regression," 179-80, *Teaching Sociology* 15 (April 1987): 178-83.

Wexler, H. K., & Lipton, D.S. (1993). From Reform to Recovery: Advances in Prison Drug Treatment. In J. Inciardi (Ed.) *Drug Treatment and Criminal Justice*. Sage Criminal Justice System Annuals, Vol. 27. California: Sage Publications, 1993.

Wexler, H.K., De Leon, G., Kressel, D., & Peters, J. (1999). The Amity prison TC evaluation: reincarceration outcomes. *Criminal Justice and Behavior*, 26, 146-167.

Wexler, H.K., Falkin, G.P., Lipton, D.S, Rosenblum, A.B., & Goodloe, L.P. (1988). A model prison rehabilitation program: an evaluation of the "Stay'n Out" therapeutic community. New York: Narcotic and Drug Research, Inc.

Winterfield, Laura. (1992). Jail Population Management Consultancy: Summary. New York: Vera Institute of Justice.

APPENDIX A:

Results of CHAID for the DOC and Combined Databases

CHAID Summary of Results for LOS Categories for DOC Database

$X^2(8) = 5807$, $p<.0001$ (Likelihood ratio, Bonferroni adjustment)
Accuracy 73%, error (27%)

Best categories to predict those released to the community (in order of strength)

11 – 16 days
17 – 25 days
26 – 36 days
102 – 133 days

With these 4 categories one can find 52 % of all the inmates released to the community in only 41 % of the sample.

Best categories to predict Prison-Bound (in order of strength)

269 – 1184 days
134 – 200 days
70 – 101 days
51 – 69 days
201 – 268 days
37 – 50 days

With these 6 categories one can find 85 % of all the Prison-Bound inmates in only 59 % of the sample.

With this database, results show that the original target of 55 days was problematic. Days leading up to 55 days are more predictive of those released to the community, but only up to 36 days. After 36 days, the next two categories (37-50, and 51-69) are weaker but more predictive of prison-bound inmates. The 102-133 day grouping can be seen in the peak on the released to the community graph, however it is the weakest grouping for the released to the community group. The second peak in the released to the community graph is overshadowed by its slightly stronger predictive strength toward the prison-bound group. This peak likely represents the 201-268 category, almost the least strongest for the prison-bound group which is likely due to a minimal pull toward the released to the community group. Overall, 11-36 days will strongly predict those released to the community, while 37-69 days will be the timeframe that has the most overlap between groups.

Using these categories, the decision was made to examine 36 days and 50 days as cutoff points in the logistic regression.

CHAID Summary of Results for LOS Categories for DOC/DCJS/CJA Database

$X^2(8) = 3334$, $p<.0001$ (Likelihood ratio, Bonferroni adjustment)
Accuracy 70%, error (30%)

Best categories to predict those released to the community (in order of strength)

- 11 – 18 days
- 19 – 30 days
- 31 – 44 days

With these 3 categories one can find 44% of the all the inmates released to the community in only 31% of the sample.

Best categories to predict Prison-Bound (in order of strength)

- 321 – 1184 days
- 172 – 239 days
- 121 – 171 days
- 61 – 88 days
- 240 – 321 days
- 89 – 120 days
- 45 – 60 days

With these 7 categories one can find 92% of all the Prison-Bound inmates in only 69% of the sample.

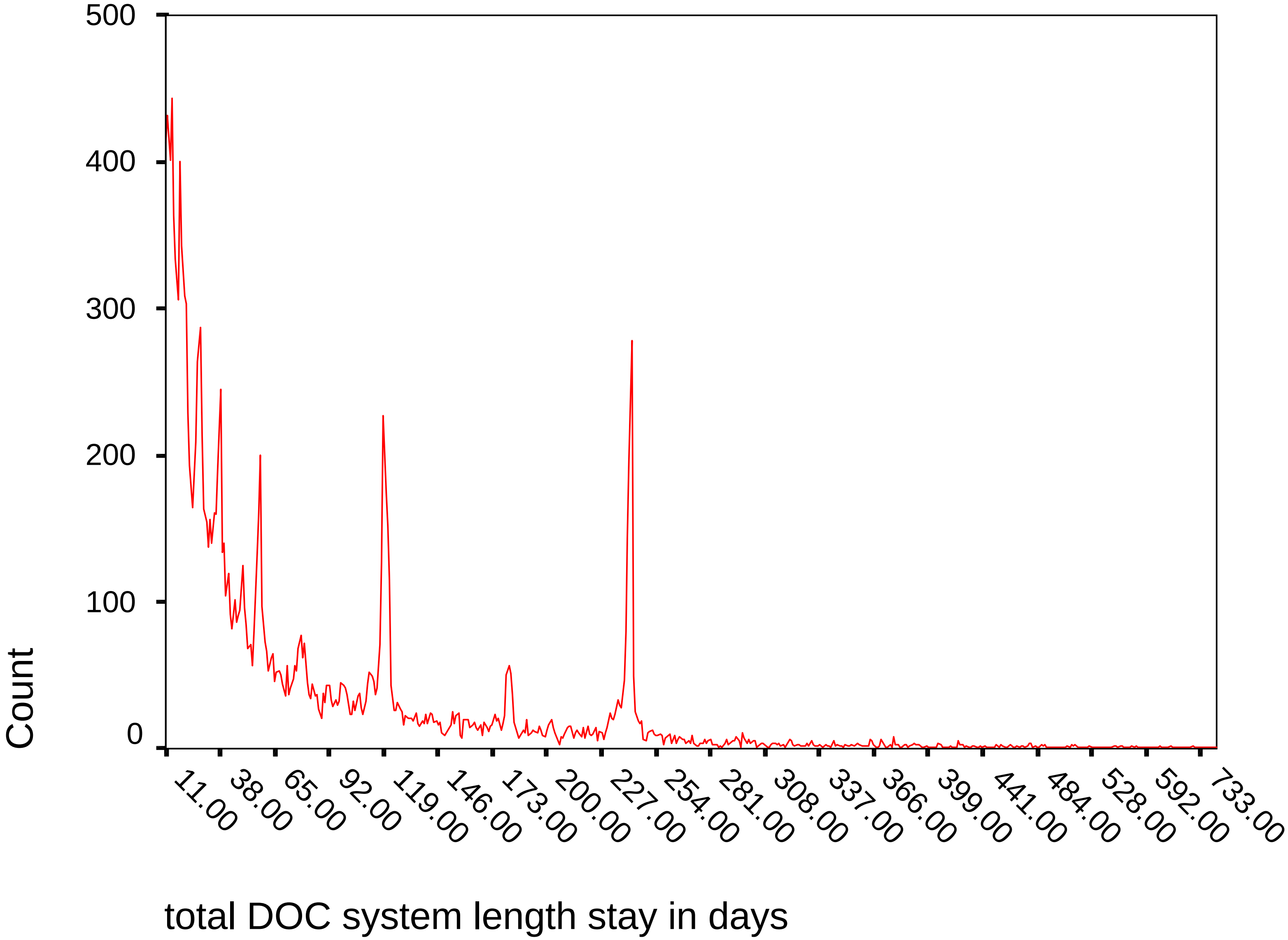
With this database, results show that the original target of 55 days was problematic. Days leading up to 55 days are more predictive of those released to the community, but only up to 44 days. After 44 days, the next category (45-60 days) is more predictive of prison-bound inmates, however it is the weakest category for the prison group. The 89-120 day grouping can be seen somewhat in the peak on the released to the community graph, however it is does not show as a predictive category for the released to the community group. The second peak in the released to the community graph also does not appear as a strong predictor for those released to the community. This peak likely represents the 240-321 category, a weaker predictor for the prison-bound group which is likely due to a minimal pull toward the released to the community group. Overall, 11-44 days will strongly predict those released to the community, while 45-60 days will be the timeframe that has the most overlap between groups.

Using these categories, the decision was made to examine 44 days and 60 days as cut-off points in the logistic regression.

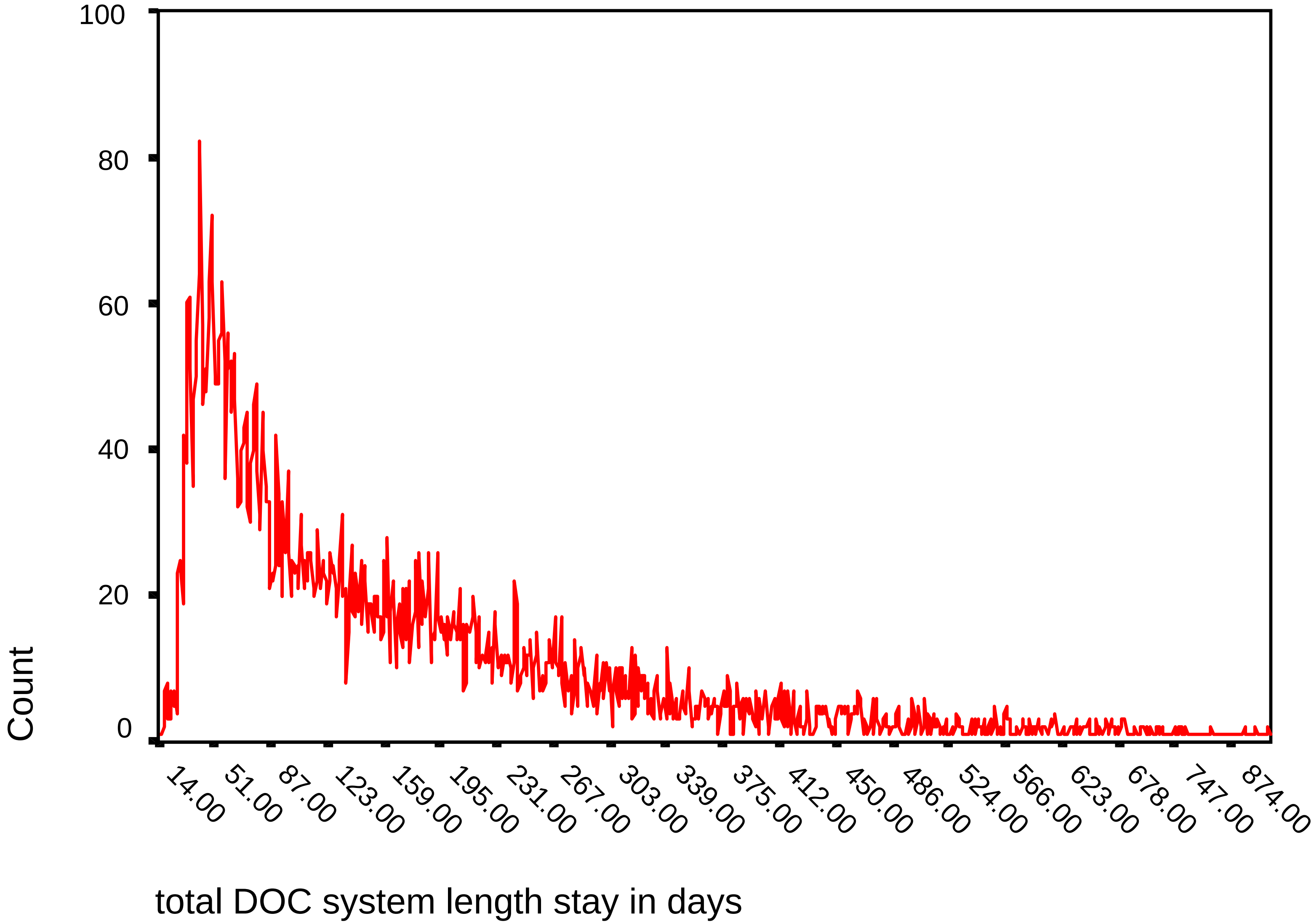
APPENDIX B:

Plots of Length of Stay (LOS) for the DOC Database

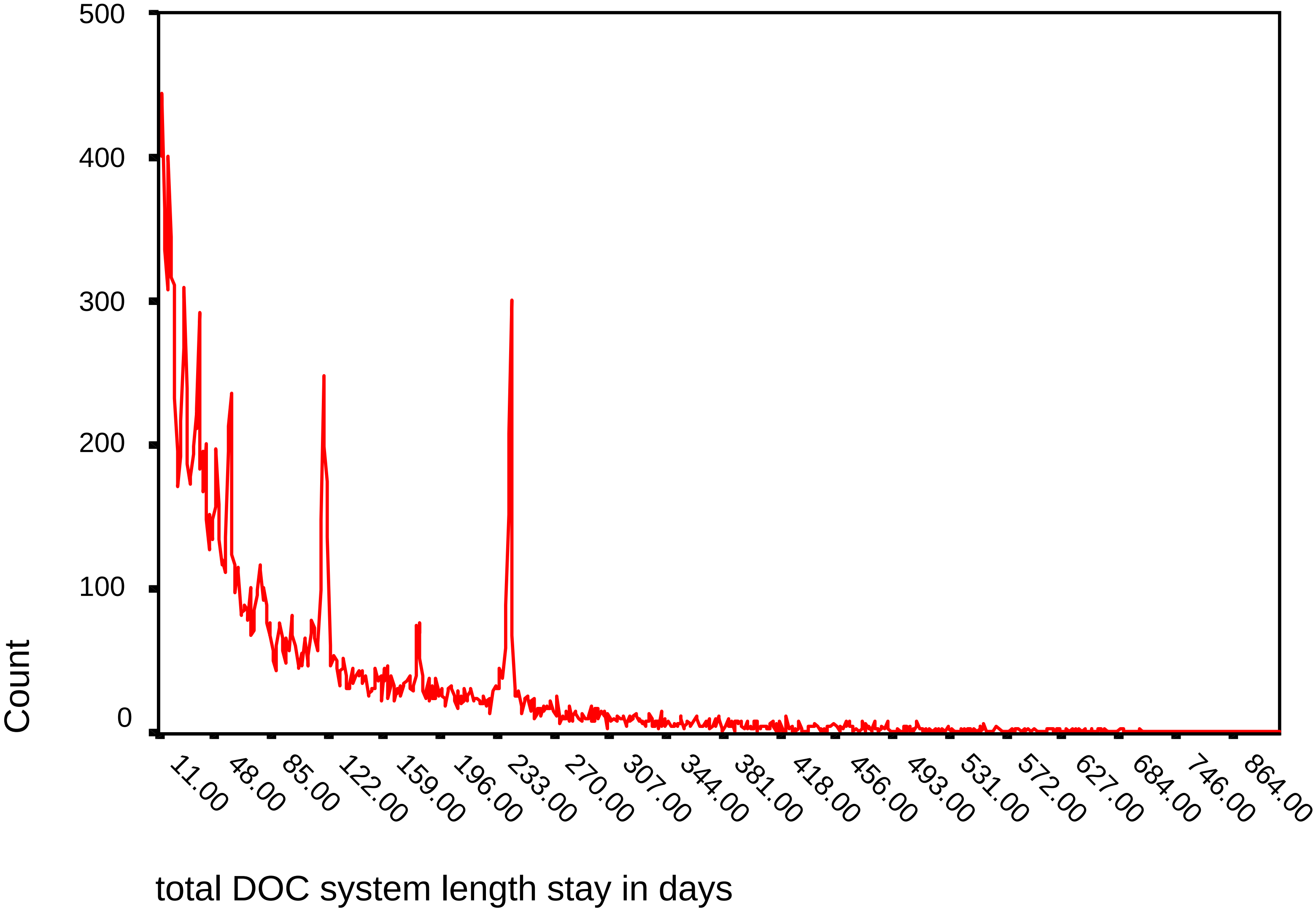
LOS FOR THOSE RELEASED TO THE COMMUNITY (n = 16,435)
DOC DATABASE



LOS FOR PRISON-BOUND (n = 7,440)
DOC DATABASE



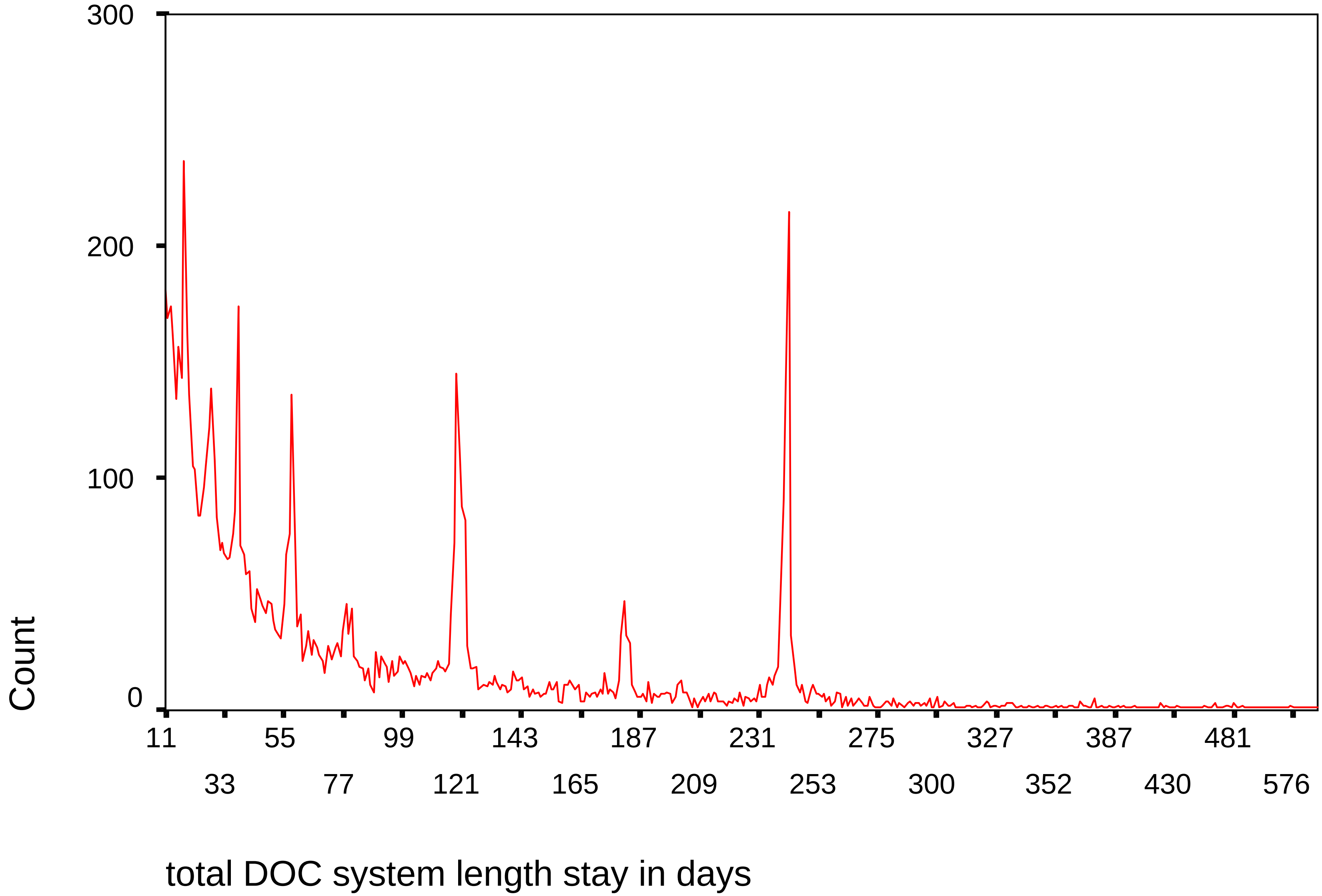
LOS ENTIRE SAMPLE (n = 23,875)
DOC DATABASE



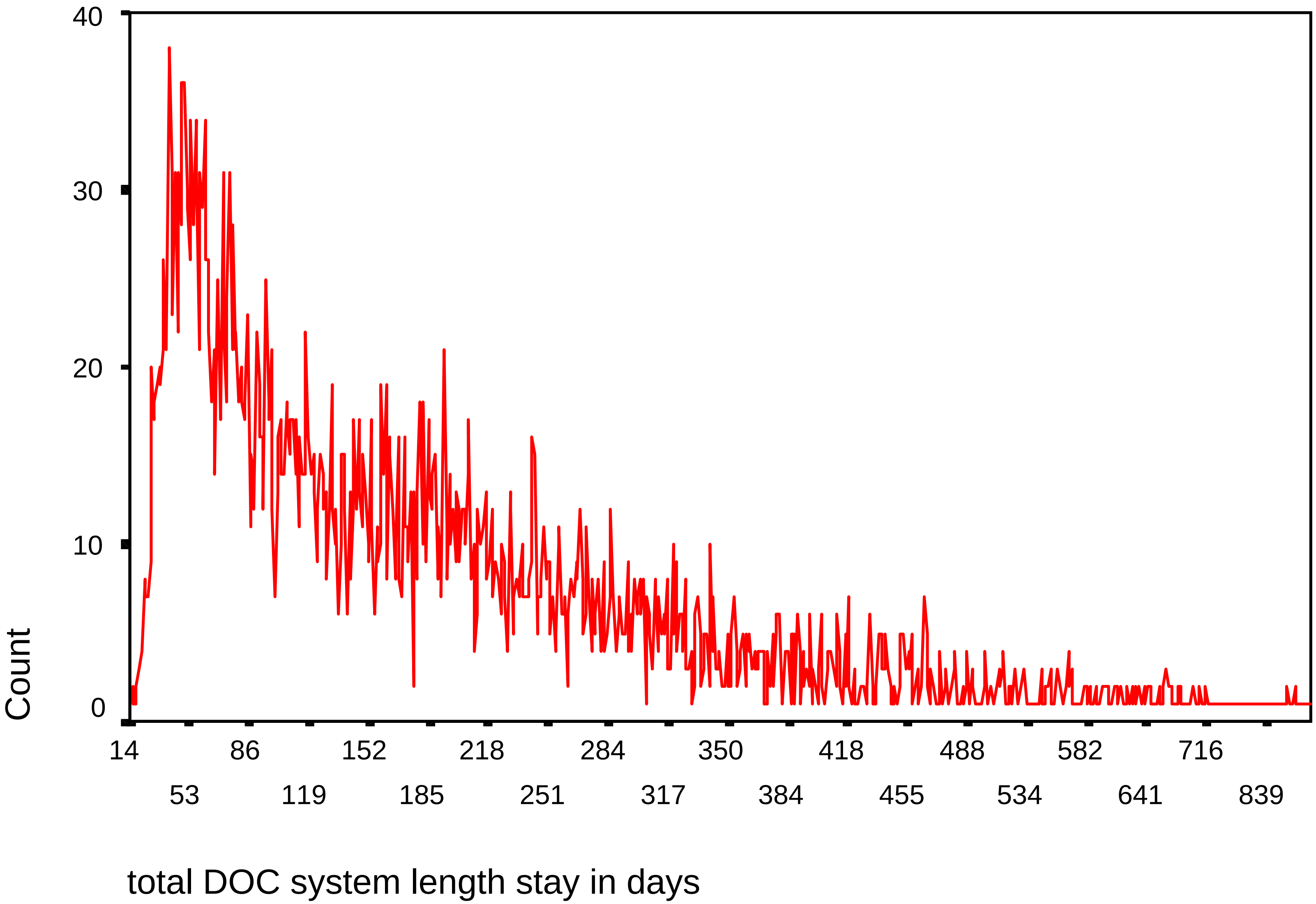
APPENDIX C:

Plots of Length of Stay for the Combined Database

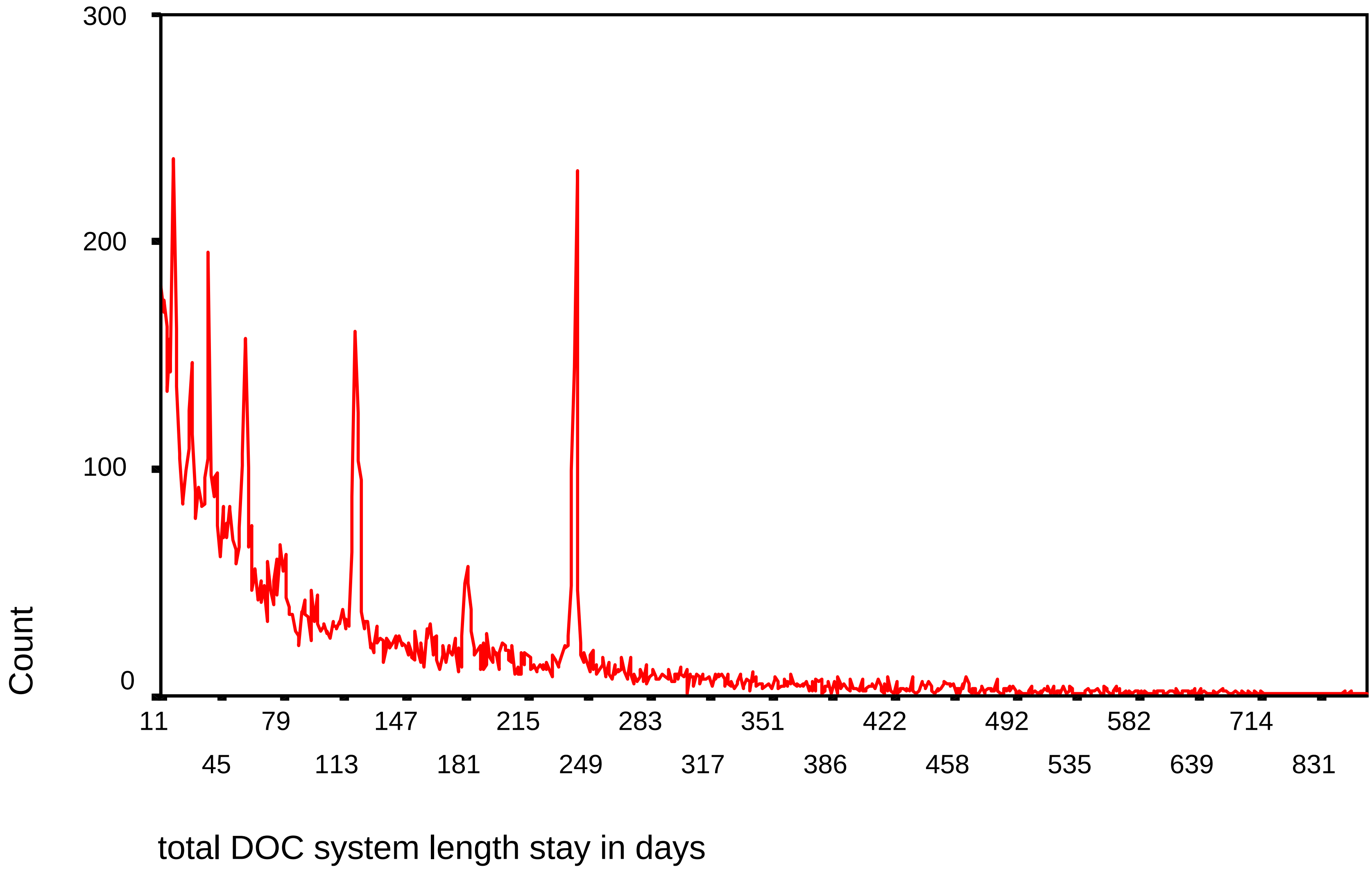
LOS FOR THOSE RELEASED TO THE COMMUNITY (n = 8,316)
DOC/DCJS/CJA DATABASE



LOS FOR PRISON-BOUND (n = 4,596)
DOC/DCJS/CJA DATABASE



LOS ENTIRE SAMPLE (n = 12,912)
DOC/DCJS/CJA DATABASE



APPENDIX D:

Guide to Implementing the Models

Implementing Logistic Regression Models

With logistic regression analysis (LRA) one predicts the probability of a dichotomous outcome (such as prison or release) or length of stay (LOS) variable. For example, LRA can be used to predict the probability that a person will stay at Rikers beyond a specified cutoff point, such as "at least 45 days." If the estimated probability is at least 50%, the inmate is classified as likely to stay at least 45 days.

The advantage of this type of model is that it is easily implemented. One need only record the person's information (e.g., DOC score, …) and multiply that information by the corresponding coefficient. Once that is done for all variables in the model, one simply adds the values. Unfortunately, this total (the logit, also known as the log odds) is not directly interpretable. One must transform this total by first exponentiating it to produce the odds that a person will stay at least 45 days. To obtain the probability estimate, one divides the odds by the sum of the odds and one. No specialized software beyond a spreadsheet would be required.

While appearing highly technical, the procedure is actually easily programmed into a spreadsheet. This is best explained with an example. Below is a step-by-step example for the logistic regression model predicting a stay at Rikers that will last AT LEAST 37 days.

1. An Information Intake Form would be created. This form would then be used on the eleventh day by the person responsible for data entry. Exhibit 1 (see the next page) is an example of a form that has yet to be completed. By default, one would assume that the person will not be released prior to 37 days (although this assumption can easily be changed and that anyone with at least a 50% chance of staying at least 37 days will be classified as staying at least 37 days.

Exhibit 1: A Sample Intake Form

Information		Put 1 Where Appropriate
Boro	Manhattan	
	Brooklyn	
	Bronx	
	Queens	
	Staten Island	
	Other Arragnment	
US Citizen	Yes	
	No	
Marital Status	Single/Never Married	
	Married	
	Living together/Common Law	
	Separated	
	Divorced	
	Widowed	
Drug Use	Yes	
	No	
Ethnicity	White/Non-Hispanic	
	Black/Non-Hispanic	
	Hispanic	
	Non-Hispanic/Other	
TopSeverity	1	
	2	
	3	
	4	
	5	
	6	
	7	
Bail	0-500	
	501-1,000	
	1,001-2,500	
	2,501-5,000	
	5,001-10,000	
	10,001-50,000	
	50,001+/Remand	
Top Remand	Yes	
	No	
Warrant	Yes	
	No	
Top Charge Drug	Yes	
	No	
Top Charge PPP	Yes	
	No	
Classification Score	Low (0-5)	
	Low Medium (6-10)	
	High Medium (11-16)	
	High (17+)	
Prior Convictions	None	
	Misdemeanor/Violation	
	Felony	

Exhibit 1, continued

Disciplinary History	Yes No	
Top Charge Violence	Yes No	
History of Violence	Yes No	
Probability of Staying at Least 37 days	0.0%	Default is 0%
Assignment (50% rule)	__	Default is No

Exhibit 2 reveals the underlying equation that would be hidden from the data entry person. This exhibit is an example of the model for a shorter length of stay when predicting the 37-day cut off.

Exhibit 2: Underlying Equation

		y/n	B	product
Boro	Manhattan	=IF(C3=1,1,0)	0	=D3*E3
	Brooklyn	=IF(C4=1,1,0)	-0.5459	=D4*E4
	Bronx	=IF(C5=1,1,0)	-0.2926	=D5*E5
	Queens	=IF(C6=1,1,0)	-0.432	=D6*E6
	Staten Island	=IF(C7=1,1,0)	0.0354	=D7*E7
	Other Arragnment	=IF(C8=1,1,0)	-0.2418	=D8*E8
US Citizen	Yes	=IF(C10=1,1,0)	0.3256	=D10*E10
	No	=IF(C11=1,1,0)	0	=D11*E11
Marital Status	Single/Never Married	=IF(C13=1,1,0)	0	=D13*E13
	Married	=IF(C14=1,1,0)	-0.1458	=D14*E14
	Living together/Common Law	=IF(C15=1,1,0)	-0.1873	=D15*E15
	Separated	=IF(C16=1,1,0)	-0.1208	=D16*E16
	Divorced	=IF(C17=1,1,0)	-0.0612	=D17*E17
	Widowed	=IF(C18=1,1,0)	-0.3078	=D18*E18
Drug Use	Yes	=IF(C20=1,1,0)	0.2078	=D20*E20
	No	=IF(C21=1,1,0)	0	=D21*E21
Ethnicity	White/Non-Hispanic	=IF(C23=1,1,0)	0	=D23*E23
	Black/Non-Hispanic	=IF(C24=1,1,0)	0.021	=D24*E24
	Hispanic	=IF(C25=1,1,0)	-0.0513	=D25*E25

. . . Continued on Next Page . . .

Exhibit 2, continued

	Non-Hispanic/Other	=IF(C26=1,1,0)	-0.1583	=D26*E26
TopSeverity	1	=IF(C28=1,1,0)	-0.1391	=D28*E28*B28
	2	=IF(C29=1,1,0)	-0.1391	=D29*E29*B29
	3	=IF(C30=1,1,0)	-0.1391	=D30*E30*B30
	4	=IF(C31=1,1,0)	-0.1391	=D31*E31*B31
	5	=IF(C32=1,1,0)	-0.1391	=D32*E32*B32
	6	=IF(C33=1,1,0)	-0.1391	=D33*E33*B33
	7	=IF(C34=1,1,0)	-0.1391	=D34*E34*B34
Bail	0-500	=IF(C36=1,1,0)	0	=D36*E36
	501-1,000	=IF(C37=1,1,0)	0.316	=D37*E37
	1,001-2,500	=IF(C38=1,1,0)	0.6	=D38*E38
	2,501-5,000	=IF(C39=1,1,0)	0.9842	=D39*E39
	5,001-10,000	=IF(C40=1,1,0)	0.7734	=D40*E40
	10,001-50,000	=IF(C41=1,1,0)	1.0472	=D41*E41
	50,001+/Remand	=IF(C42=1,1,0)	1.5579	=D42*E42
Top Remand	Yes	=IF(C44=1,1,0)	0.0703	=D44*E44
	No	=IF(C45=1,1,0)	0	=D45*E45
Warrant	Yes	=IF(C47=1,1,0)	1.1805	=D47*E47
	No	=IF(C48=1,1,0)	0	=D48*E48
Top Charge Drug	Yes	=IF(C50=1,1,0)	-0.1977	=D50*E50
	No	=IF(C51=1,1,0)	0	=D51*E51
Top Charge PPP	Yes	=IF(C53=1,1,0)	-0.0129	=D53*E53
	No	=IF(C54=1,1,0)	0	=D54*E54
Classification Score	Low (0-5)	=IF(C56=1,1,0)	0	=D56*E56
	Low Medium (6-10)	=IF(C57=1,1,0)	0.0722	=D57*E57
	High Medium (11-16)	=IF(C58=1,1,0)	0.0767	=D58*E58
	High (17+)	=IF(C59=1,1,0)	2.3274	=D59*E59
Prior Convictions	None	=IF(C61=1,1,0)	0	=D61*E61
	Misdemeanor/Violation	=IF(C62=1,1,0)	0.2833	=D62*E62
	Felony	=IF(C63=1,1,0)	0.2363	=D63*E63
Disciplinary History	Yes	=IF(C65=1,1,0)	0.3631	=D65*E65
	No	=IF(C66=1,1,0)	0	=D66*E66
Top Charge Violence	Yes	=IF(C68=1,1,0)	-0.1079	=D68*E68
	No	=IF(C69=1,1,0)	0	=D69*E69
History of Violence	Yes	=IF(C71=1,1,0)	0.0895	=D71*E71
	No	=IF(C72=1,1,0)	0	=D72*E72
			constant	1.2767
			logit	=SUM(F3:F74)
			odds	=EXP(F76)
Probability of Staying at Least 37 days	=F78		probability	=F77/(F77+1)
Assignment (50% rule)	=IF(B78>=0.5,"Yes","No")			

2. Exhibit 3 shows the default values (0) for all the data elements. The default probability estimate when all values are the default, as shown at the bottom of the exhibit, is 0.7819.

Exhibit 3: Default Values for all Data Elements

Information	Put 1 Where Appropriate y/n	B	Product
Boro	Manhattan	0	0.0000	0.0000
	Brooklyn	0	-0.5459	0.0000
	Bronx	0	-0.2926	0.0000
	Queens	0	-0.4320	0.0000
	Staten Island	0	0.0354	0.0000
	Other Arragnment	0	-0.2418	0.0000
US Citizen	Yes	0	0.3256	0.0000
	No	0	0.0000	0.0000
Marital Status	Single/Never Married	0	0.0000	0.0000
	Married	0	-0.1458	0.0000
	Living together/Common Law	0	-0.1873	0.0000
	Separated	0	-0.1208	0.0000
	Divorced	0	-0.0612	0.0000
	Widowed	0	-0.3078	0.0000
Drug Use	Yes	0	0.2078	0.0000
	No	0	0.0000	0.0000
Ethnicity	White/Non-Hispanic	0	0.0000	0.0000
	Black/Non-Hispanic	0	0.0210	0.0000
	Hispanic	0	-0.0513	0.0000
	Non-Hispanic/Other	0	-0.1583	0.0000
TopSeverity	1	0	-0.1391	0.0000
	2	0	-0.1391	0.0000
	3	0	-0.1391	0.0000
	4	0	-0.1391	0.0000
	5	0	-0.1391	0.0000
	6	0	-0.1391	0.0000
	7	0	-0.1391	0.0000

Exhibit 3, continued

Bail	0-500	0	0.0000	0.0000
	501-1,000	0	0.3160	0.0000
	1,001-2,500	0	0.6000	0.0000
	2,501-5,000	0	0.9842	0.0000
	5,001-10,000	0	0.7734	0.0000
	10,001-50,000	0	1.0472	0.0000
	50,001+/Remand	0	1.5579	0.0000
Top Remand	Yes	0	0.0703	0.0000
	No	0	0.0000	0.0000
Warrant	Yes	0	1.1805	0.0000
	No	0	0.0000	0.0000
Top Charge Drug	Yes	0	-0.1977	0.0000
	No	0	0.0000	0.0000
Top Charge PPP	Yes	0	-0.0129	0.0000
	No	0	0.0000	0.0000
Classification Score	Low (0-5)	0	0.0000	0.0000
	Low Medium (6-10)	0	0.0722	0.0000
	High Medium (11-16)	0	0.0767	0.0000
	High (17+)	0	2.3274	0.0000
Prior Convictions	None	0	0.0000	0.0000
	Misdemeanor/Violation	0	0.2833	0.0000
	Felony	0	0.2363	0.0000
Disciplinary History	Yes	0	0.3631	0.0000
	No	0	0.0000	0.0000
Top Charge Violence	Yes	0	-0.1079	0.0000
	No	0	0.0000	0.0000
History of Violence	Yes	0	0.0895	0.0000
	No	0	0.0000	0.0000
			Constant	1.2767
			Logit	1.2767
			Odds	3.5848
Probability of Staying at Least 37 days	78.2%		Probability	0.7819
Assignment (50% rule)	Yes			

3. Exhibit 4 is an example of a completed form for a fictitious inmate. It is assumed that the person:

- was arraigned in Manhattan;
- is NOT a U.S. citizen;
- is widowed;
- has no self reported drug use;
- for whom the top severity is a misdemeanor B;
- for whom bail was 0 to 500 dollars;
- was not remanded;
- for whom there was no warrant;
- the top charge is a drug charge and a crime against person/property
- has no prior convictions;
- has not prior violence or disciplinary history; and
- for whom the top charge is not for a violent crime.

For this example, the estimated probability that the person will stay at Rikers at least 37 days is 40.8%, which does not reach the 50% threshold. Therefore, one would score the inmate as unlikely to stay at least 37 days.

Exhibit 4: Completed Form for a Fictitious Inmate

Information		Put 1 Where Appropriate	y/n	B	product
Boro	Manhattan	1	1	0.0000	0.0000
	Brooklyn		0	-0.5459	0.0000
	Bronx		0	-0.2926	0.0000
	Queens		0	-0.4320	0.0000
	Staten Island		0	0.0354	0.0000
	Other Arragnment		0	-0.2418	0.0000
US Citizen	Yes		0	0.3256	0.0000
	No	1	1	0.0000	0.0000
Marital Status	Single/Never Married		0	0.0000	0.0000
	Married		0	-0.1458	0.0000
	Living together/Common Law		0	-0.1873	0.0000
	Separated		0	-0.1208	0.0000
	Divorced		0	-0.0612	0.0000
	Widowed	1	1	-0.3078	-0.3078
Drug Use	Yes		0	0.2078	0.0000
	No	1	1	0.0000	0.0000
Ethnicity	White/Non-Hispanic		0	0.0000	0.0000
	Black/Non-Hispanic		0	0.0210	0.0000
	Hispanic		0	-0.0513	0.0000
	Non-Hispanic/Other	1	1	-0.1583	-0.1583

Continued on next page….

Exhibit 4, continued

TopSeverity	1		0	-0.1391	0.0000
	2		0	-0.1391	0.0000
	3		0	-0.1391	0.0000
	4		0	-0.1391	0.0000
	5		0	-0.1391	0.0000
	6		0	-0.1391	0.0000
	7	1	1	-0.1391	-0.9737
Bail	0-500	1	1	0.0000	0.0000
	501-1,000		0	0.3160	0.0000
	1,001-2,500		0	0.6000	0.0000
	2,501-5,000		0	0.9842	0.0000
	5,001-10,000		0	0.7734	0.0000
	10,001-50,000		0	1.0472	0.0000
	50,001+/Remand		0	1.5579	0.0000
Top Remand	Yes		0	0.0703	0.0000
	No	1	1	0.0000	0.0000
Warrant	Yes		0	1.1805	0.0000
	No	1	1	0.0000	0.0000
Top Charge Drug	Yes	1	1	-0.1977	-0.1977
	No		0	0.0000	0.0000
Top Charge PPP	Yes	1	1	-0.0129	-0.0129
	No		0	0.0000	0.0000
Classification Score	Low (0-5)	1	1	0.0000	0.0000
	Low Medium (6-10)		0	0.0722	0.0000
	High Medium (11-16)		0	0.0767	0.0000
	High (17+)		0	2.3274	0.0000
Prior Convictions	None	1	1	0.0000	0.0000
	Misdemeanor/Violation		0	0.2833	0.0000
	Felony		0	0.2363	0.0000
Disciplinary History	Yes		0	0.3631	0.0000
	No	1	1	0.0000	0.0000
Top Charge Violence	Yes		0	-0.1079	0.0000
	No	1	1	0.0000	0.0000
History of Violence	Yes		0	0.0895	0.0000
	No	1	1	0.0000	0.0000
				constant	1.2767
				logit	-0.3737
				odds	0.6882
Probability of Staying at Least 37 days	40.8%			probability	0.4076
Assignment (50% rule)	__				

Implementing the Survival Analysis

Survival analysis offers greater flexibility than logistic regression in that it predicts days at Rikers and these predicted values can easily be translated to various cutoffs. A drawback is that the models are not as readily scored as logistic regression models because the distributional assumption underlying the model must also be incorporated. One can program the scoring algorithm into an application. If this approach is taken, one can score inmates on a daily basis using the program.

Another strategy is to rely on existing statistical software, e.g., STATA or SAS. The approach would require that a database be constructed. It must include the data set used to construct the survival analysis models used in this report. One must also add a field that indicates whether the case was in the model development database. As new inmates are to be scored, their information must be appended to this data set. The indicator must identify these cases as NOT being in the development database. One can then re-estimate the model using the development sample, but also score the new cases.

Exhibit 5 (on the next page) displays an example layout of the required database. The first 16,435 cases would be those used to build the model. Each would have a Yes flag indicating that it was in the model development sample. The data for all other subsequent inmates would be appended to the end of this file (starting with sequence number 16,436), and the flag would indicate that those data were not part of the development sample. One would then need re-run the survival analysis selecting only those cases in the development sample to estimate (in this case re-estimate) the model, but scoring all the cases.

This strategy requires the necessary software, the proper maintenance of the inmate database, and writing the code necessary for model re-estimation and whole data set scoring.

NOTE about Exhibit 5: For the sake of space and readability, this exhibit vertically stacks blocks of data that, in a real application, would appear side-to-side. That is, <u>all</u> information about a particular case (as in sequence, or case, number 16,435 below) would appear on a single line across to ensure that data relevant to a given person will be correctly associated with that person. The sample data are shown as they are below because in the format of this research document they would not fit horizontally across the page.

Exhibit 5: Example Layout of the Database for a Survival Analysis

Sequence #	Development Sample	Manhattan	Bronx	Queens	Staten Island	Other Arrangements	Male	US Citizen	Single, Never Married	Drug Use
1	Y	Y	N	N	N	N	Y	Y	Y	N
2	Y									
.	Y									
.	Y									
.	Y									
16,435	Y									
16,436	N									
16,437	N									
16,438	N									
16,439	N									
16,440	N									

African-American/Non-Hispanic	Felony A	Felony B	Felony C	Felony D	Felony E	Missing Top Severity	Bail 501-1000	Bail 1001-2500	Bail 2500-5000	Bail 5001-10000
N	N	N	Y	N	N	N	N	N	N	N

Bail 10001-50000	Bail 50001+	Warrant Present	Violent	Drug	Other	Classification Score	Score > 15	Misde-meanor Prior
N	Y	Y	Y	N	N	14	N	N

Continued on the next page

Exhibit 5, continued

Felony Prior	Remanded	Disciplinary History	Violenc e History	Age (years)
Y	N	N	N	25