RECIDIVISM REPORT: Technical Appendix

DATA COLLECTION

Definition of Recidivism

The Department identifies a recidivist as an inmate who, after release from prison, commits a new offense that results in a commitment to the Department. This definition, focusing on new offenses committed as opposed to return to supervision or custody alone, is intended to cast recidivism more in the broader scope of public safety costs rather than in the narrower scope of corrections system costs.

Offenders who return to prison for a technical violation of supervision are not counted as recidivists unless they committed a new offense. Offenders do count as recidivists for a new offense committed while on supervision even when, as often happens, the conviction for the offense occurs after they return to prison for technical violations.

The Department defines the standard recidivism rate at two years after prison release. This follow-up period optimizes the standard rate as a tool for evaluating both sentencing policies and prison programs and interventions intended to reduce recidivism. Two years allows enough time from release for recidivism rates to become stable and reliable measurements. Also, rates within two years of release are more likely than those from a longer follow-up period to reflect any effects of incarceration on recidivism.

Studies typically use one or more of three recidivism measures: rearrest, reconviction, and recommitment to prison. Each of these measures has strengths and weaknesses. Arrest data is the broadest measure of crime available, but an arrest does not imply that a new offense actually occurred. Conviction data indicates that a new offense did occur, but does not necessarily indicate the seriousness of the offense. Commitment data is the narrowest measure, but does indicate that a new offense of a serious nature did occur.

The Department only receives data on those offenders who are recommitted for either a term of supervision or imprisonment and, therefore, can only reliably report the rate of recidivism based on new offense data drawn from recommitment records.

In order to provide further insight into recidivism of inmates released from the Department, data on arrests have been collected. This data has been used to calculate rearrest of released inmates as a secondary measure of recidivism. This secondary measure is reported because many studies use rearrest, rather than reconviction, as their only recidivism measure.

Data Sources: Release Cohort and Reoffense Data

Data for this report were extracted from the Department of Corrections' Offender Based Information System (OBIS) for the Bureau of Research and Data Analysis. The data used represents inmate releases, used to define the inmate release cohort analyzed, and sentence component information, used to identify offenses for which inmates were convicted subsequent to release from prison.

The inmate release data contain one record per inmate released from July 1, 1993 through June 30, 2000. Demographic information (e.g., age, sex, race) and commitment data (e.g., primary offense type, time served in prison, custody level at release) is collected from release records. Records from this data form the cohort of inmates analyzed for this report. Only inmates released permanently are included—that is, the cohort includes all inmates whose incarceration sentence has been satisfied. This includes some inmates whose sentence involves a period of post-prison supervision.

In previous editions of this report, offense data was collected from data on admissions to the Department for community supervision or imprisonment. For this edition, published in May 2001, offense data was obtained only from a single source, consisting of all sentence components. This data includes all component information, including offense dates, for every offense for which an offender was sentenced to the Department. This data is used to identify the dates on which an inmate has committed offenses.

Data Sources: Rearrest Data

The Florida Department of Law Enforcement (FDLE) provided arrest data for this report. FDLE maintains the Florida Criminal Information Center (FCIC) which receives arrest reports from local police agencies within the state. Since arrest reports from local agencies are not mandated by law, this data may underreport actual arrests of released inmates. Computerized criminal history files drawn from this statewide database were used to provide arrest data to the Department of Corrections.

A data file of 107,280 inmates in the release cohort identified for this report was provided to FDLE, including two personal identifiers: the FDLE number and the Social Security number, as recorded in the Department of Corrections data base. Matches were found for a total of 83,394 offenders.

The Florida Department of Law Enforcement provided two sets of data. The first included matched based on FDLE number. The second included additional matched based on Social Security number. Each data file contained the arrest date for every arrest reported to FDLE. A total of 588,270 arrests were for inmates in the release cohort subsequent to their release from prison.

Case Selection

In order to get the most accurate recidivism rate possible given available data sources, a number of release cases are not considered in the recidivism analysis. Inmates who are released to another state are omitted from study. Since the inmate is living in a different state, his chance of returning to Florida's corrections system is fairly low. Including those inmates would lower these recidivism rates. Data sources are not currently available to the Department to allow analysis of these cases.

Some inmates released from prison to community supervision technically violate the terms of their supervision and return to prison to be released again later. Only the first release is kept in the current analysis. Since these multiple releases are occurring with no intervening crime being committed, including them would over-count the number of "successes." The exception to this is the case where the inmate is returned to prison on a technical violation, but is subsequently sentenced while in prison for a crime committed between his prior release and his readmission for the technical violation. In that case, the next release following the new offense is also included.

Another issue that may affect recidivism rate calculations is the time, often lengthy, from when an offense occurs to the date the offender is readmitted to the Department. Since the data on the offense is not available until the inmate is actually readmitted to the Department (at which time the data entry is done), the data collection must allow a sufficient amount of time for the offender to be arrested, tried, sentenced, and returned to the Department's custody or supervision. The delay between when an offense is committed and when the offender reenters the Department’s information system must be taken into account to prevent this data collection time from biasing results of the analysis.

Of inmates readmitted to the Department for new offenses, the distribution of time between the new offense date and the subsequent admission date indicates that 75 percent of inmates are received within 250 days. Therefore, inmate releases less than 250 days prior to data collection for this report are not included in the analysis. Releases on or after January 27, 2000 were therefore excluded from this analysis.

Using this "buffer" period alleviates the problem of considering a released inmate a success (non-recidivist) when, in fact, he has not had sufficient time to return to the Department. For example, for an inmate released on January 27, 2000, data collection would have to wait until October 4, 2000, (250 days after release) in order to achieve a 75 percent probability of capturing that the inmate committed a new offense on the day he was released. Since offense data available for this analysis is only current through October 3, 2000, this release record would be removed from consideration.

The final release cohort included 120,881 release cases. This includes 107,255 inmates, of which 13,626 had more than one permanent release during the period.

DATA ANALYSIS

Recidivism Rate Estimation Method

Two vital data elements are needed for each released inmate: the follow-up period and the time to reoffense. The follow-up period is defined as the number of months between the date of release and the date 250 days prior to the data collection date (October 3, 2000 for this report). Every released inmate has a follow-up period. The time to reoffense is defined as the number of months from the date of release to the date a new offense was committed. A non-recidivists will obviously not have a time to reoffense.

These two values are used to determine whether the inmate *failed* (recidivated) or was *censored* (did not recidivate during the follow-up period), and the time before each event occurred. For example, an inmate who was released and committed a new offense 6 months later is represented as having *failed at time 6 months*. An inmate who was released and did not reoffend in the 15 months for which data is available, is represented as being *censored at time 15 months*. The designation of an inmate as being censored means only that he has not recidivated up to a certain point in time—he may or may not recidivate in the future.

The Kaplan-Meier product limit method of estimating the recidivism function is a technique used in survival analysis to incorporate both the failed and censored records. Basically, it uses the conditional probability of not failing ("surviving") during each time period, given that the inmate had not failed before that time period. The product of these conditional probabilities is the "survival distribution function". The complement of the survival distribution function is the recidivism function represented in this report.

The following is an example of how the product limit method is calculated. The table below shows a cohort of inmate releases and the events (failures and censoring) that take place at each time period. The mathematical formulae for computing the recidivism function follows the table.

Months Since Release	Inmates Surviving at Beginning of Time Period = n(t)	Failures During Time Period = f(t)	Inmates Censored During Time Period = c(t)	Inmates Surviving at End of Time Period	Conditional Probability of Survival = p(t)	Survival Function = S(t)	Recidivism Function = 1 - S(t)
0	100	0	0	100	1.000	1.000	0.000
1	100	5	3	92	0.950	0.950	0.050
2	92	6	4	82	0.935	0.867	0.133
3	82	10	2	70	0.878	0.761	0.239
4	70	15	10	45	0.786	0.598	0.402

The conditional probability of surviving through period (t), given that the inmate still survives at beginning of period (t), is:

$$p(t) = \frac{n(t) - f(t)}{n(t)} \quad \text{and} \quad n(t) = n(t-1) - f(t-1) - c(t-1)$$

where *d(t)* = number of inmates that fail during period (t)
n(t) = number of inmates alive at beginning of period (t)

The survival function, or probability of surviving, through period (t) is the product of the conditional probabilities of surviving all preceding time periods, including the current period (t):

$$S(t) = \prod_{k \leq t} p(k)$$

where *t* = the current time period
p = conditional probability of surviving a time period
k = all time periods for which *p* has been calculated

The recidivism function, or failure rate, is the complement of the survival function:

$$1 - S(t)$$

An important property of this estimation method is that, if there are no censored observations, it returns a recidivism rate as calculated in the more conventional way:

$$\text{Recidivism rate at time (t)} = \frac{\text{total number of recidivists at time (t)}}{\text{total number of releases}}$$

In other words, when all releases have the same follow-up period, the recidivism rate is simply the percentage calculated by dividing the number of recidivists by the number of total number of releases.

Using this method allows more data to be used in describing the recidivism rates over time. It also allows for statistical procedures that can be used to test for differences in the recidivism function between different groups of inmates.

The same survival analysis technique has been used to estimate the rearrest rate of released inmates. That analysis was conducted in the same method as the reoffense rate estimation, but with the first arrest after the prison release substituted for the first offense. To allow comparison of rearrest and reoffense rates, the same prison release events used for the reoffense measurement were also used to measure the rearrest rates.

Selection of Factors Related to Recidivism Rates

Scientific studies have documented that many factors contribute to recidivism rates. These include such personal characteristics as age, gender, race, length and severity of criminal history, education and skill levels, and conduct in prison. Other factors such as length of incarceration are believed to affect recidivism rates as well. In order to use recidivism rates as performance measures or evaluation criteria, it is important to know how these factors affect recidivism rates. Therefore, certain factors known to influence recidivism rates have been analyzed for this report.

The factors reported here were selected using the following criteria, which should ensure that future reports, analyses, and evaluations of Department programs and activities can depend on this report's data, sources, and methods:

- strong evidence from major studies showing the factor influences recidivism rates
- the factor is measured independently from inmate's participation in specific programs or services the Department provides
- valid data are readily available from the Department's information system to measure the factor
- data obtained from Department sources are expected to remain reliable over time.

Factor data were grouped into categories primarily to facilitate presentation of the basic relationship each has to recidivism rates. Rather than using a rigorous statistical technique to establish the groups, the groups were based on a combination of these criteria:

- inmate sub-populations for which data is often requested
- inmate groupings typically reported in other Department documents
- equal distribution of the cases, where possible.

These criteria for categorizing the factors serve two goals: to show each factor's basic relationship, if any, to recidivism rates; and to provide data that is needed to respond to typical internal and external information requests the Department receives. These categorizations need not be used in future analyses or evaluations, although this report will continue them in order to provide data continuity to support overall trend identification and analysis in the future.

Combined Effect of Factors on Recidivism Rates

Among the factors analyzed, the following nine factors are significantly related to recidivism: age at release, prior recidivism, race, gender, education level, offense type, custody level at release, number of disciplinary reports received, and time served in prison. Additional factors analyzed which did not maintain a significant relationship to recidivism against these nine factors included: psychological diagnosis severity, percentage of sentence served, and type of disciplinary reports received. Certain other factors, including Hispanic ethnicity and post-release formal supervision were not analyzed. Available data is not sufficient to analyze pre-admission and post-release factors such as employment and drug and alcohol use.

Even without these additional variables or more intensive factor measurements, the nine significant factors identified do improve our ability to predict whether or not released inmates will recidivate. A logistic regression model was used to determine how well these combined factors reduce the expected error in prediction at three follow-up periods after release from prison. Since logistic regression models require a dichotomous dependent variable (e.g., recidivate or non-recidivate), the model was tested at more than one point along the cumulative distribution function (recidivism rate curve) to determine how much the model's predictive power varied with the follow-up period analyzed.

To test the model's predictive power, a cross validation method was used in which the data set was randomly divided in half to create a prediction sample and a validation sample. The data set used for the analysis included 109,971 release cases. Releases for the partial fiscal year from July 1999 through January 2000 were excluded. Of these, 87,917 cases (80.0%) had no missing values on any factor variables included in the model and could be used in the analysis. An analysis of cases with missing data revealed no apparent systematic relationship of missing data to the recidivism/non-recidivism dichotomy.

From the prediction sample a logistic regression model generated the recidivism likelihood based on the nine factors. The model's recidivism probabilities were applied to the validation sample in order to compare whether inmates predicted to recidivate actually did so. The statistical model was tested at three follow-up periods (18, 36, and 60 months after release) to ensure that the model's predictive value does not vary unacceptably across follow-up intervals.

The method for calculating this improvement in prediction was as follows, using the 36 months rate analysis as an example. Five factor measures where dichotomized as follows: gender (male/non-male), race (black/non-black), prior recidivism (yes/no), release custody classification (minimum-medium/other), violent primary offense (yes/no), and property primary offense (yes/no). Note that although primary offense type, whether violent or property, was incorporated as two separate identity variables in the model, they reflect two aspects of the same factor. Since the remaining four factor measures—age, education level, total disciplinary reports, and time in prison—are continuous rather than categorical variables, the natural logs of their values were used. These continuous variables were categorized and tested for nonlinearity without problematic results, although both time in prison and education level may have somewhat nonlinear relationships to recidivism.

A stepwise logistic regression analysis of this model was applied to the prediction sample data set yielding results presented in the table below.

Logistic Regression Results for Predictive Random Sample				
Factor Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	2.1682	0.1721	158.7	<.0001
Age	-0.7779	0.0459	287.4	<.0001
Disciplinary reports	0.2623	0.0172	233.7	<.0001
Education level	-0.1629	0.0204	63.9	<.0001
Gender	0.2460	0.0382	41.5	<.0001
Months in prison	0.1218	0.0192	40.3	<.0001
Prior recidivism	1.1689	0.0441	703.3	<.0001
Property offense	0.1889	0.0274	47.5	<.0001
Race	0.4495	0.0249	326.7	<.0001
Release custody	-0.2270	0.0381	35.5	<.0001
Violent Offense	-0.5060	0.0292	300.1	<.0001

These parameter estimates were then applied to the factor values for the cases in the validation sample to obtain an odds value. The formula for applying the estimates is:

$$odds = \exp\begin{pmatrix} \boldsymbol{b}_0 + \boldsymbol{b}_1 PRIOR + \boldsymbol{b}_2 AGELOG + \boldsymbol{b}_3 VOFF + \boldsymbol{b}_4 BLACK + \\ \boldsymbol{b}_5 MALE + \boldsymbol{b}_6 DRLOG + \boldsymbol{b}_7 RLCLASS + \boldsymbol{b}_8 TIMELOG + \\ \boldsymbol{b}_9 SCORELOG + \boldsymbol{b}_{10} POFF \end{pmatrix}$$

where	PRIOR=1	if prior recidivist, 0 otherwise
	AGELOG	is natural log of age (in years) at release
	VOFF=1	if violent primary offense, 0 otherwise
	BLACK=1	if inmate is black, 0 otherwise
	MALE=1	if inmate is male, 0 if female
	DRLOG	is natural log of number of disciplinary reports incurred
	RLCLASS=1	if custody is minimum or medium, 0 if close
	TIMELOG	is natural log of time in prison in months
	SCORELOG	is natural log of latest education score
	POFF=1	if property primary offense, 0 otherwise

The odds value was converted to the probability an inmates would recidivate:

$$prob = \frac{odds}{(1 + odds)}$$

This decimal probability was rounded to the nearest thousandth. Cases with probability values greater than .500 were defined as predicted to recidivate by the model, and those with probabilities less than .500 were defined as predicted not to recidivate. A few cases (103) with values equal to .500 were classified as unpredicted.

By forcing the model to make predictions in this manner, it is possible to compare whether cases predicted to recidivate and not to recidivate with whether the case actually recidivated or not in the validation sample. The procedure for calculating this is described by the following table.

Logistic Regression Model Prediction Performance					
Total Cases				35,929	
Total Cases Correctly Predicted				22,090	61.5%
	Failure	11,872	53.7%		
	Success	10,218	46.3%		
Total Cases Misclassified				13,839	38.5%
	Predicted failure	7,008	50.6%		
	Predicted success	6,728	48.6%		
	Unclassified failure	53	0.4%		
	Unclassified success	50	0.4%		
Chance Error Probability		p (1-v) + v (1-p) =			49.9%
		Failures	N	Failures / N	Complement
Prediction sample (p)		18,619	36,010	0.517	0.483
Validation sample (v)		18,653	35,929	0.519	0.481
Analysis of Model Prediction					
	Errors	% of N	Correct	% of N	
Chance	17941	49.9%	17988	50.1%	
Model	13839	38.5%	22090	61.5%	
Percent Reduction in Errors (PRE)		(chance errors – model errors) / chance errors			= 22.9%
Percent Improvement in Prediction (PIP)		(model correct – chance correct) / chance correct			= 22.8%

The percent reduction in error (PRE) calculation above is a standard method for assessing the predictive value of a logistic regression model. The conversion of this into a percent improvement in prediction (PIP) is an atypical, but more easily communicated result that is analogous to the PRE.

The ability of the model to predict recidivism better than chance is significantly reduced if the prediction decision is not constrained to the .500 probability value. The table below demonstrates that as the decision criterion is relaxed, the number of model errors increases and the PRE declines. However, this is also clearly due to the substantial increases in unpredicted cases that appear with the less stringent criteria.

Logistic Regression Prediction Decision Criterion Analysis

		Correctly Predicted		Incorrectly Predicted			
Decision Criterion	Model Errors	Actual Failure	Actual Success	Predicted Failure	Predicted Success	Unclassified	PRE
0.500	13,839	11,872	10,218	7,008	6,728	103	22.9%
0.505	14,363	11,600	9,966	6,766	6,484	1,113	19.9%
0.510	14,886	11,342	9,701	6,510	6,230	2,146	17.0%
0.515	15,412	11,057	9,460	6,241	5,987	3,184	14.1%

The table below indicates how well these factors predict recidivism in our inmate release population for three follow-up periods. For example, consider the model's performance at 36 months. From the probability of recidivating based on the prediction and validation samples, we would expect 17,941 errors and 17,988 correct predictions by chance alone—that is, knowing nothing about the inmates in the validation sample other than when they were released and when they reoffended. However, when using these nine factors, the statistical model allows only 13,839 errors and makes 22,090 correct predictions, increasing our percentage predicted correctly from the expected 50.1% to an actual 61.5%. In other words, knowing these nine factors our ability to predict whether or not an inmate will recidivate by 36 months after release is improved by 22.8%.

	18 Months After Release	36 Months After Release	60 Months After Release
Prediction Sample Size	43,877	36,010	27,683
Validation Sample Size	44,040	35,929	27,821
Errors Expected by Chance	19,433	17,941	10,949
Correct Predictions Expected	24,607	17,988	16,872
Prediction Errors by Model	14,769	13,839	7,238
Correct Predictions by Model	29,271	22,090	20,583
Expected Percent Correct	55.9%	50.1%	60.6%
Actual Percent Correct	66.5%	61.5%	74.0%
Prediction Improvement	19.0%	22.8%	22.0%

This report demonstrates that certain factors influence recidivism rates for inmates released from Florida state prison facilities. These data support certain findings of other studies that show, for example, that younger inmates recidivate sooner and at higher rates than older inmates do. Establishing the combined effect of these factors does not address the issue of their relative influence on recidivism rates—that is, determining which factors are more strongly related to recidivism than others are. However, it is clear that these factors should be considered when conducting evaluations using recidivism rates as performance measures.

Relative Effect of Factors on Recidivism Rates

To determine the relative influence of these factors, we focus not on whether inmates recidivate, but on how long it takes for inmates to recidivate. This requires a statistical technique different from logistic regression. Logistic regression is suitable for analyzing dichotomous dependent variables, such as whether or not an inmate recidivates. But for that analysis, it is necessary to limit the follow-up period to a certain time during which it is known whether or not an inmate recidivated. Further, cases without that minimum follow-up (i.e., censored cases) must be excluded from the analysis to avoid biasing the results. In addition, all information regarding those excluded cases and how their associated factors contribute to recidivism is lost.

These problems of artificially fixing a follow-up period and ignoring censored case data are overcome by analyzing the timing of events—recidivism in this case—which survival analysis is essentially designed to do. Recidivism data, in fact, are ideal for conducting survival analysis precisely because the time origin required by the analysis is clearly defined (i.e., by an exit from prison in this study).

Four alternative regression models available for the analysis were considered for this study: log-normal, exponential, Weibull, and log-logistic—which are defined by their respective distributions of time to recidivism. The recidivism hazard function is an inverted "U" shape skewed sharply positive. Both an extreme log-normal and certain log-logistic models theoretically describe this better than either the exponential or Weibull models, because the former approach "0" rather than infinity at the time origin.

However, with the large number of cases available for the study, each of these models performs similarly in describing the data with the factors included. This appears to be a result of the extraordinary sharpness of the hazard function in the total release sample. The partial recidivism rate vaults from 0 to its peak (2.69%) at month three after release, and then drops precipitously to less than half that rate (1.28%) by month 11, to less than a third (.85%) by month 18, and to less than a fourth (.65%) by month 24. The underlying hazard function may be too skewed in a large sample to allow meaningful discrimination between models that begin at the origin and those descending from infinity.

Since our task is merely to identify and describe the relative effects of recidivism factors, three criteria were used to select a preferred model. The first is a standard evaluation using the best log likelihood statistic generated by the model. The second is a comparison of how well the log likelihood statistic is improved from each unfitted model when fitting the factor regressors. This is measured by a generalized R^2 statistic, where the higher value indicates greater improvement of the log likelihood. The third criterion is which fitted model yields regression parameters with the lowest standard errors. The table below presents these criteria values for each model.

Survival Analysis Regression Model Selection Criteria				
Citerion	Log-Normal	Log-Logistic	Weibull	Exponential
Log L statistic	-114177	-115323	-116758	-118894
Log L reduction fitted	.29	.29	.30	.31
Standard errors:				
Maximum	.03184	.03173	.03021	.02302
Average	.0169	.0168	.0158	.0121

Based on the simple log likelihood statistic criterion, the log-normal followed by the log-logistic models would be preferred. However, the reduction in log likelihood statistic criterion when the model is fitted to the regressors indicates the most improvement is achieved with the exponential followed by the Weibull models. The exponential model, a special case of the Weibull, was rejected because the hazard function proves not to be constant over time. Finally, the lowest standard errors for the factors are achieved using the Weibull model.

Another basis on which to select the regression model for the survival analysis is to compare the direction and magnitude of each factor's estimated parameter across the models. No differences in direction (i.e., whether the factor increased or decreased time to recidivism) were found for any factor, regardless of the model analyzed. However, two important minor differences were detected between the models regarding the magnitude of factors.

First, only the log-logistic model found a significant, small independent effect of an identity variable for primary drug offense with the property offense variable in the model. In the other models, any effect of this variable was eliminated when the primary property offense identity variable was included. Since this effect only appears in the log-logistic model, it is reasonable to conclude that the effect—for the entire prison release population—is an artifact of the model rather than a real effect of the factor.

Second, only the log-normal and log-logistic models found a significant, albeit tiny, effect for the time in prison factor across the entire release population. At the same time, these models shifted the property offense factor one place in an ordering of the factor parameter magnitudes, by raising that value slightly. The logistic regression analysis showed that the time in prison factor has a nonlinear relationship to recidivism—appearing significant for the analysis at certain follow-up periods but not at others. It is possible that the small relationship appearing in the log-normal and log-logistic models for time to recidivism is inappropriately measured using these simple models.

This suggests that an interaction effect between time in prison, offense type, and possibly other factors would model the effect of time in prison more appropriately. This interaction effect is likely to appear significant either or both for a limited portion of the hazard function and for a subset of the total release population. These features of the different models argue for testing several models when conducting evaluation studies using recidivism as a performance measure, especially for inmate subsets that are not statistically representative of the total release population.

For this report, proportional hazards regression—a technique suitable for analyzing a Weibull model—was selected to analyze each factor's independent effect on recidivism while controlling for the effects of other significant factors. This statistical procedure reveals three things about a particular factor:

- whether it has a statistically significant effect on recidivism,
- whether it accounts for more variation in recidivism rates than others,
- whether and by how much it tends to raise or lower recidivism rates.

This model of the time to recidivism, measured in months, yielded the results presented in the table below. This model included all factors determined to be significant in the logistic regression in analyzing whether or not an inmates recidivated.

Proportional Hazards Regression Results:
Estimating Time to Recidivism in Months

Factor Variable	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
Age	-0.0212	0.0007	1047.8	<.0001
Disciplinary reports	0.0163	0.0009	352.5	<.0001
Education level	-0.0295	0.0018	281.4	<.0001
Gender	0.2168	0.0182	141.4	<.0001
Months in prison	-0.0015	0.0003	34.1	<.0001
Prior recidivism	0.4583	0.0159	832.4	<.0001
Property offense	0.1304	0.0121	117.1	<.0001
Race	0.3618	0.0119	927.7	<.0001
Release custody	-0.1148	0.0162	50.0	<.0001
Violent offense	-0.3847	0.0136	800.5	<.0001

A proportional hazards regression allows a reasonably straightforward interpretation of the factor parameters based on the hazard ratios deriving from them. The interpretation is, however, different for factors according to whether they are measured as categorical and continuous variables. The table below provides the hazard ratio and its raw interpretation value for each factor.

Proportional Hazards Regression Results: Interpretation Based on Hazard Ratios			
Factor Variable	Variable Type	Hazard Ratio	Raw Interpretation
Age	Continuous	0.979	-2.1%
Disciplinary reports	Continuous	1.016	1.6%
Education level	Continuous	0.971	-2.9%
Months in prison	Continuous	0.999	-0.1%
Gender	Categorical	1.242	24.2%
Prior recidivism	Categorical	1.581	58.1%
Property offense	Categorical	1.139	13.9%
Race	Categorical	1.436	43.6%
Release custody	Categorical	0.892	-10.8%
Violent offense	Categorical	0.681	-31.9%

For categorical factors the interpretation is a relative hazard that inmates with the characteristic will recidivate as compared with inmates not having that characteristic. A hazard ratio of 1.0 would mean no difference in this hazard. Factors with a value more than 1.0 are more likely to recidivate; those with a value less than 1.0 are less likely. The percentage difference in hazard is the ratio's difference from 1.0. For example, inmates who have recidivated at least once before are 58.1% (1.581 – 1.0) more likely to reoffend than those who have not.

For continuous factors the interpretation is similar but the increased or decreased hazard must be evaluated per unit of value on the factor. To obtain this estimated percent change in the hazard per unit change in the factor's value, multiply by 100 the absolute value difference between the hazard ratio and 1.0. For example, an inmate's hazard for recidivating decreases by 2.1% for each year older the inmate is at release. These interpretations must be understood as true of the entire cohort of inmates analyzed rather than as true for any particular inmate in the sample.

To determine the relative impact of these factors on time to recidivism, a method must be used to weight the continuous factor measures against the categorical measures. A reasonable way to do this is to multiply the continuous factor's hazard ratio by the mean of the factor. The table below provides these average effects for the continuous factors.

Order of Factor's Influence on Time to Recidivism Based Weighted Values from Hazard Ratios			
Factor Variable	Raw Interpretation	Factor Mean	Average Effect
Age	-2.1%	32.0	-67%
Prior recidivism	58.1%	na	58%
Race	43.6%	na	44%
Violent offense	-31.9%	na	-32%
Gender	24.2%	na	24%
Education level	-2.9%	7.0	-20%
Property offense	13.9%	na	14%
Release custody	-10.8%	na	-11%
Disciplinary reports	1.6%	5.1	8%
Months in prison	-0.1%	32.2	-3%

Discussion of Specific Factors Influencing Recidivism Rates

This section discusses each factor included in the report, how the factor was measured, how it was grouped for reporting purposes, and what the factor may truly represent. Note that the groups established for reporting purposes are not those used in any of the regression analyses conducted.

Release Year

The release year reported here is the fiscal year for state government in Florida, which extends from July 1 to June 30. Inmate release dates are grouped according to these annual time periods. Breaking the releases down by this factor allows one to note the development of any trends in recidivism rates, which may appear. None appear in the data available for this report. This factor alone is not expected to show a relationship to recidivism rates.

However, if an upward or downward trend appears in this factor in the future, the Department will examine what other factors this trend may reflect, including changes in release population demographics, offender supervision practices, judicial behavior on recommitment of technical violators, law enforcement practices, etc. This release-year analysis also satisfies information requests for the latest recidivism rates available.

Age

Release age is calculated as the inmate's actual release date minus the reported date of birth. Inmates released more than once in the data have different release dates accordingly. The age groups reported were devised primarily to achieve similar numbers of cases with smaller categories at the younger and older extremes to meet information request needs.

However, this does not mean age in itself, understood as maturity level, contributes to recidivism rates. Other factors inherent in age may actually form the basis for the relationship. For example, older inmates may recidivate at lower levels simply because they have less mobility and energy to commit aggressive action which many crimes require. Or younger inmates may simply have less experience which leads either to a greater propensity to commit crimes or a lesser ability to conceal crimes committed. Evaluators using recidivism rates must control for the effects of age, but may also have to examine what it is about age, if anything particular, that contributes to recidivism rates in order to properly understand the effect of a program.

Gender and Race

The vast majority of inmates self-report as white or black, so the race data are simplified into three categories: white, black, and other. Race is categorized simultaneously with gender only to facilitate reporting here. The number of females in the "other" race category is so small in the release population since 1993 that reliable estimates of that group's recidivism rates is not yet possible.

Since males typically account for more criminal offenses, commit offenses more frequently, and commit more serious offenses than females, it is not unexpected that males recidivate at higher rates. The relationship between race and recidivism rates is less clear. For example, white males appear to recidivate less than black males, and white females recidivate less than black females. However, the difference in rates on the race factor among females is substantially less than the difference between race among males. This suggests that some particular aspect of the race and gender combination underlies the relationship between race and recidivism rates.

Education Level

The education grade reported here is the result of Adult Basic Education tests used to assess inmates while in prison. The groups used are those consistent with other Department documents reporting this information.

These test scores are used instead of self-reported education level data for two reasons. First, the tests provide a more-reliable measure of an inmate's true educational level. Second, inmates' scores on these tests may change while in prison due to participation in education programs. The latest test score prior to release is used as an indicator of the inmate's level of education at release, whether or not an inmate received any educational programming in prison. Since this is a diagnostic tool to help determine an inmate's need for education while incarcerated, more inmates are assessed using these tests than those who take education courses. Therefore, this data constitutes a measure of the inmate's education level independent from any program participation. Evaluations of the recidivism effect of education programs while in prison will require an analysis of the effect of these courses on an inmate's subsequent test scores.

Prior Recidivism

Prior recidivism is measured by identifying the number of releases counted for recidivism purposes prior to the most recent release. Since the data available only include releases that occurred after 1992, this factor currently undercounts the extent of prior recidivism events—that is, offenses committed after releases that occurred before 1993. The effect of this factor may change in future years as more data becomes available and if other reliable measures of criminal history are developed.

Offense Type

The offense type categories used—violent, property, and drug—are common among studies and reports using criminal justice data. Released inmates are classified based on the most serious offense for which they were imprisoned. The most serious offense is defined as that which, first, had the longest sentence imposed and, second, the highest felony degree.

Offenses included in the violent category for this report are those which meet the following standard definition by the Florida Department of Corrections. A crime is defined as violent if the crime involves actual or the threat of physical harm to a person or the crime has a reasonable probability of causing unintended physical harm or physical threat of harm to a person. One of the following conditions must occur for a crime to be defined as violent under this definition:

- Actual physical harm.
- Actual threat of physical harm.
- A reasonable probability existed that individual criminal acts could have resulted in unintended physical harm.
- A reasonable probability existed that individual criminal acts could have resulted in the threat of physical harm.

Time Served in Prison

The number of months actually spent in prison were grouped into categories primarily to achieve similar numbers of cases. The amount of time spent in prison is determined by several factors including sentence length, behavior in prison, whether an inmate qualifies for early-release programs, etc. Therefore, it is expected that the effect of time in prison on recidivism rates will vary depending on how these other underlying factors vary.

For example, longer stays in prison appear to reduce recidivism rates, but this may be because violent offenders, who recidivate at lower rates, tend to stay in prison longer. In addition, the longer the stay in prison, the greater the inmate's age is when released, which also tends to lower recidivism rates. Also, there is some evidence that there is an interaction between the time in prison and other factors such as primary offense type. More complex multivariate models are required to specify nature of the effect length of imprisonment has on recidivism rates.

Custody Class

The groups used reported here are the three primary custody levels: close, medium, and minimum. Maximum custody inmates include only those few who are under a death sentence. In future reports, a new lower level of custody ("community") will have to be analyzed, but this has yet to appear in our release cohort.

An inmate's custody class depends on several factors including: seriousness of offense, sentence length, time to release, behavior in prison, work history in prison, participation in programs, etc. Most inmates' custody levels have been reduced to medium or minimum security levels prior to their release. Inmates still in close custody at release tend to be more serious offenders and more violent and disruptive inmates. Program participation is more available for medium and minimum custody inmates. Evaluations using recidivism rate measures should take into account whether inmates included in the evaluation were in close custody at release.

Total Disciplinary Reports

The number of disciplinary reports received in prison typically indicates the amount of rule-breaking and criminal behavior offenders commit in prison. The groups reported here are designed primarily to capture certain types of inmates. Many inmates receive no disciplinary reports, and many others receive only a few during their incarcerations. Other inmates receive substantial numbers of disciplinary reports throughout their incarcerations.

The number of disciplinary reports received could be a function of other factors including: aggressiveness, physical or medical condition, unwillingness to abide by rules, or simply the length of imprisonment. Nevertheless, the factor is somewhat predictive of recidivism rates after release.

Persistence of the Factors' Relationships to Recidivism Rates

In 1989, the Bureau of Justice Statistics, an agency of the U.S. Department of Justice, published a Special Report, Recidivism of Prisoners Released in 1983. This study examined more than 108,000 inmates released in 1983 from 11 large states, including Florida, whose reoffenses were tracked for three years. That study examined several factors related to recidivism rates and documents relationships similar to those reported here for factors.

These two studies used different data sources, measurement methods, and analysis techniques and covered two very different geographic areas and prison release cohorts. The Bureau of Justice Statistics (BJS) study included prison releases from 11 states, whereas this study included releases only in Florida. The BJS study included inmates released across state lines and counts reoffenses regardless of which state they occur in, whereas this study only included inmates released to Florida who reoffend in this state. Finally, the BJS cohort were inmates released in one year (1983), whereas this study includes inmates released during six and one-half years (July 1993 through January 2001). Both studies analyzed very large samples of inmate releases.

The point of comparing these two studies is two demonstrate the persistence of the relationship between certain factors and recidivism. Factors analyzed in both studies included:

- age
- gender
- race
- education
- offense type
- time in prison.

The influence these factors have on recidivism rates remains notable across regions and time periods, even using such different data sources and analysis methods. This demonstrates the general stability and robust quality of the relationships between many of these factors and recidivism rates.

At a minimum, therefore, evaluation studies of inmates released from prison should take these common factors into account if they use recidivism rates as performance measures. In addition, other factors included in this study should also be analyzed in such evaluations. Recidivism rate outcomes for prison inmates are not meaningful unless they take these relationships into account and control for their effects.

In late 2000 or early 2001, BJS will likely publish the first results of a more recent recidivism study. This new study of state prison inmates released in 1994 is similar methodologically to the study of inmates released in 1983. This should allow an assessment of changes in rates over a decade. More importantly it will likely continue to support the importance of these factors and perhaps others to prison inmate recidivism.

Relationship of Factors to Recidivism Rates in Two Studies: Percentage Reincarcerated within 3 Years of Release

		BJS Report: 1983 releases		FDC Report: 1993 – 2000 releases
		From 11 States	From Florida	In Florida
Total Releases		108,580	13,105	120,881
Age	**Under 18**	50.6 %	*	55.8 %
	18 – 24	44.9 %	44.5 %	45.7 %
	25 – 29	43.2 %	36.6 %	43.0 %
	30 – 34	43.0 %	37.3 %	42.3 %
	35 – 39	36.5 %	26.9 %	38.6 %
	40 – 44	30.7 %	24.6 %	35.1 %
	45 or older	25.7 %	22.9 %	22.5 %
Gender	**Male**	41.9 %	38.1 %	40.8 %
	Female	33.0 %	26.9 %	34.8 %
Race	**White**	38.0 %	31.1 %	31.8 %
	Black	45.3 %	45.9 %	46.2 %
Education	**Under 9th grade**	38.4 %	40.6 %	42.7 %
	Some high school	40.9 %	44.0 %	36.2 %
	High school graduate	35.0 %	29.7 %	31.0 %
	Some college or more	30.4 %	21.5 %	
Offense	**Violent**	36.5 %	34.5 %	31.0 %
	Property	47.7 %	43.1 %	45.8 %
	Drug	30.3 %	22.9 %	44.3 %
Time Served in Prison (months)	**Under 6**	**	42.4 %	35.7 %
	7 – 12	**	34.9 %	41.1 %
	13 – 18	**	35.8 %	41.3 %
	19 – 24	**	36.3 %	42.3 %
	25 – 36	**	34.5 %	41.8 %
	37 or more	**	32.2 %	37.7 %
Total Population Rate		41.4 %	37.6 %	40.2 %

* Too few cases to provide a reliable estimate.
** Data not available.

Recidivism rate values should not be compared across studies that differ to such degree in sources and methods. For example, it would be inappropriate to compare results from the Florida sample in the BJS study to rates published in this report to claim that overall recidivism rates have either declined slightly or not changed much over a decade.

ADVANTAGES AND LIMITATIONS OF THE METHODOLOGY

This methodology for calculating recidivism rates is an improvement over earlier methods used by the Florida Department of Corrections in several ways. Specific advantages and limitations of this methodology are detailed below.

Case Selection

In the method only the first release is counted for inmates who have returned to prison multiple times for technical violations of a supervision sentence. This methodology effectively eliminates the multiple "successes" for inmates that would return and be released from prison many times, with no new crime being committed.

Many offenders return to prison for technical violations of post-release supervision conditions such as a new felony arrest which results later, upon conviction, in a new offense date. Earlier methods could not use this new offense date because the offenders had already returned to prison on their original sentences. The method counts this subsequent offense date.

Data Sources

The method uses all offense information in the Department's Offender Based Information System (OBIS) to determine whether and when a new offense was committed. In the past, only specific types of inmate movements into the system were available to determine that an inmate was a recidivist.

The method is, however, restricted in the amount of historical data available. The Department's improved release data begins in January 1993, whereas the former method used data from 1988. This will affect such things as the prior recidivism factor and the correction for multiple releases, because missing data makes measurement of these past events incomplete. As the amount of data available grows in the future, the effects of these limitations will become better defined.

Analysis Technique

The methodology has major advantages over former methods that will make recidivism rates more useful as management and evaluation tools. Former methods used the more conventional way of calculating recidivism rates. The entire cohort of releases was restricted to only those cases that had at least a 24-month follow-up period. In addition, a 6 month "buffer" period was used so that the entire group had ample time to appear in the Department's data system. Therefore, to calculate a two-year recidivism rate on inmates released during a particular year, data could not be collected until more than 30 months after the end of the release period. Thus the data used was necessarily old by the time it could be collected, analyzed, and reported.

This method uses survival analysis techniques to estimate the recidivism function at each month following release. The method can use relatively new release information in calculating the rates. This may allow new trends in recidivism rates to be identified, analyzed, and reported much more quickly and more reliably.

This method allows recidivism rates to be computed over time instead of only at a fixed follow-up period. Therefore, the recidivism function can be graphed by time-to-reoffense to show different rates of increase and when the rates begin to "level off." This allows evaluators to compare results to valid benchmark rates at whichever follow-up period is most appropriate for the program or group under study.

This method allows more data to be used in calculating rates. The level of influence demographic and other factors have on recidivism rates can be demonstrated better by using larger amounts of data. The method counts recidivists and non-recidivists as such during their appropriate follow-up period even when less than 24 months. The greater amount of release and reoffense data available, the more clearly and accurately any effect of programs and other factors can be demonstrated.

This analytical method allows the direction and magnitude recidivism rates to be measured, regardless of the follow-up periods available for each released inmate. As such it creates a framework within which any effect on recidivism programs and other activities may have on recidivism rates. The analysis can be conducted using the maximum amount of data available. In addition, the framework is flexible enough to allow evaluation of both large and small scale programs and to assess both short-term and long-term effects.

Method Limitations

The current methodology has some limitations. These are considered subjects for future improvement in the reporting of general recidivism rates. In addition, methods for overcoming these limitations may be attempted in conducting any evaluations using this approach.

First, some released inmates return to prison for technical violations of supervision conditions and remain in prison for a considerable time. Although inmates can and do commit new offenses against correctional officers, other inmates, and prison property, the prison environment greatly reduces the likelihood that they will commit new offenses for which they will be formally charged and convicted. Thus, these technical violators returned to prison could be considered either "not at risk" or "at low risk" of reoffending while in prison during their follow-up period.

Such inmates are considered "successes" (non-recidivists) for this report. It is unclear whether and to what extent inclusion or exclusion of these inmates from this study would alter these recidivism rates. The Department's view is that these inmates should not be excluded from the analysis for the same reason that inmates who reoffend with minor crimes that result in local jail time or local supervision can not be excluded. The point of this report is to document the relationship between factors outside the control of the corrections system on recidivism rates of state prisoners released. Future studies of recidivism and program evaluations using recidivism measures should examine the extent to which local corrections systems and returning technical violators to state prison prevent released inmates from reoffending.

Another limitation of this method is that inmates released out-of-state are removed from the analysis. This is because the Department does not currently have access to corrections data from any state except Florida. Using only Florida data to determine whether an inmate who was released out-of-state reoffended and returned to the Department's custody would skew the recidivism rates lower.

Finally, the analytic framework this method provides may be of limited use in conducting evaluations of programs or activities involving small numbers of inmates. Analysis of this data clearly shows that for very small subsamples of inmates, recidivism rates estimated using survival analysis techniques might not be sufficiently stable to provide reliable results, while controlling for important factors related to recidivism. The minimum population of inmates for which this framework can be properly applied has yet to be established.

REFERENCES

Allison, Paul D. (1995) Survival Analysis Using the SAS® System: A Practical Guide. Cary, NC: SAS Institute Inc.

Beck, Allen J. and Bernard D. Shipley (1989) Recidivism of Prisoners Released in 1983. U.S. Department of Justice, Bureau of Justice Statistics, Special Report.

DeMaris, Alfred (1992) Logit Modeling: Practical Applications. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-086, Newbury Park, CA: Sage Publications Inc.

London, Dick (1988) Survival Models and Their Estimation. Winsted, CT: ACTEX Publications.

Schmidt, Peter and Anne D. Witte (1988) Predicting Recidivism Using Survival Models. New York: Springer-Verlag.