Aging in Prison: Arguing for Elderly Compassionate Release Programs

Ozdemir Erdemir and Harsh Dedhiya; in collaboration with SPARK and Jenifer McKim

Project Description

The number of prisoners age 55 and older sentenced to more than one year in prison has jumped 400 percent over the last two decades, leaving hundreds of thousands of elderly behind bars. Yet some states, including Massachusetts, still don't have compassionate release programs to help seniors who may no longer be a security risk. In this project, we attempt to show that the growing number of elderly prisoners is a matter of concern to the state of Massachusetts, and that they are not as large of a threat to society when compared to prisoners of other ages. We then determine if there are conditions under which elderly prisoners currently are favored to get conditional releases, such as early releases with parole due to good behavior. Lastly, we highlight conditions under which elderly prisoners become recidivists, or enter prison again after being released once.

We aim to predict the number of elderly prisoners that get conditional releases based upon race, committed crime, and level of education in the future in Massachusetts. These variables were selected as race and level of education are the two biggest features that define the demographics of our data set. Looking to see if these variables have an effect on whether a judge allows a conditional release could show that some demographics have biases against them during this process. Committed crime is similar in that it defines how dangerous a prisoner is, and this can result in changes in the judge's decision making process. Looking at these three features may highlight trends that are not obvious at first glance. Predicting the number of elderly prisoners that get conditional releases would first involve predicting the overall size of the prison population in Massachusetts with a least-squares polynomial fit, and then performing a logistic regression on the independent variables to determine which ones are more important. The odds developed by the regression should allow us to predict the number of people with a set of characteristics that get conditional releases. It is essential that states are able to predict how many elderly are in jail in the future, and that states are equipped to deal with the rising number of seniors.

We also wanted to discover how race and offense affect the chances of an elderly individual becoming a recidivist, or a criminal who is incarcerated more than once, in Massachusetts. This would also involve a logistic regression on those independent variables. Learning more about factors that affect rates of recidivism may underscore demographics in which increased social services and awareness would lower crime rates for elderly individuals. A deep-dive analysis of these problems with the help of Spark would help NECIR, the New England Center for Investigative Reporting, in writing impactful stories about vulnerable peoples and, hopefully, lead to national policy changes, like informing policy concerning compassionate release programs in Massachusetts.

Data Description

Originally we had intended to gain access to the National Corrections Reporting Program's (NCRP) restricted data, which is administered by the Bureau of Justice Statistics, and archived with the Inter-university Consortium for Political and Social Research (ICPSR). This data would have been constructed and sent to us with more detailed information that we would have needed to answer our original proposal questions, namely the way the elderly die in prison. We were in contact with NCRP for a month before it was revealed that we would need IRB approval to be granted access to the data. Because IRB approval is beyond the scope of the class, we were forced to modify our original proposal questions and instead try to find meaningful correlations within the more limited, but public, NCRP prison dataset.

We used four separate public data sets from NCRP, which is internally labelled as ICPSR 36404. The Term Records dataset has one record for each separate term in prison, as well as an ID assigned to each record. The Prison Admissions dataset contains one record for each admission to prison, and a prisoner will have more than one record if they were admitted more than once. The Prison Releases dataset contains one record for each release from prison, and a prisoner will have more than one record if they were released more than once. The last dataset is the Year-End population dataset, which contains one record for each prisoner who is in custody on December 31st for each year. All of the four datasets contain similar personal information for each inmate: Sex, Race, Education, Age, Crime, State, and Admission/ Release information, but only the first dataset has IDs for each inmate.

One of the main difficulties with the NCRP data is that all of the data is in bins: instead of being given a prisoner's age of admission we would only know which category the prisoner falls into for age of admission (18-24, 25-34, …., 55+). This is true for almost all of the numerical components of our dataset. Due to this categorical nature of our data, we found it appropriate to use Logistic Regression to explore our data.

We make considerable use of the release dataset. Specifically, we use the type of release prisoners receive for our regression analysis. There are 4 types of release: "Conditional release", "Unconditional release", "Other(including death, transfer, AWOL, escape)", and "Missing". Conditional release refers to prisoners who are allowed to leave prison before their sentence due to good behavior and proving they are no longer a threat to society. An unconditional release signifies that the prisoner got released regularly. When we analyze our data using these categories, we do not consider prisoners with "Other" or "Missing" values.

The NCRP public datasets are large. Each dataset has between 10-18 million entries, and after dropping entries with missing values and cutting the dataset down into relevant categories we want to study (Age: 55+ prisoners, State: Massachusetts), we were still left with 20,000 entries.

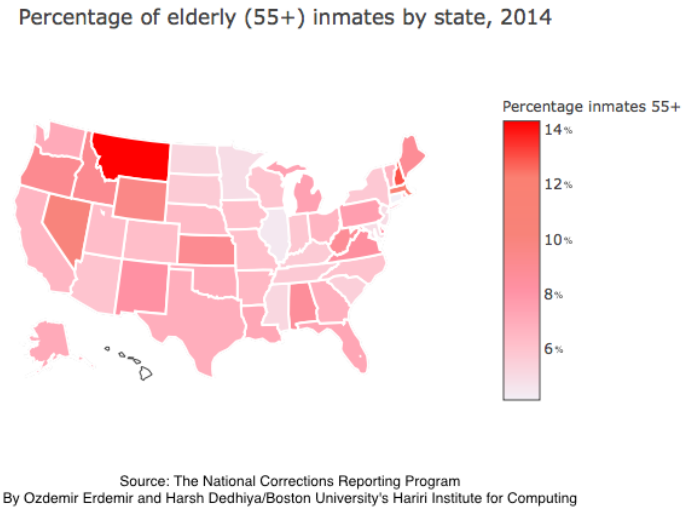

Figure 1. Showing the states with the highest percentage of elderly inmates. The top 3 states are Montana (14%), New Hampshire (12.9%) and Massachusetts (7.2%)

Data Analysis

Initially we started by gathering some basic statistics about the percent of elderly people in prison around the country and in Massachusetts. Surprisingly, we found that in 2014, the percent of elderly people in Massachusetts compared to the overall population in prison was the 3rd highest in the country when looking at the Term Records set (Figure 1). Thus, questions and inquiries related to the elderly prisoner population are more significant locally in Massachusetts.

Later, we also check the percentage of conditional releases in Massachusetts among elderly prisoner releases, which shows that the percentage is relatively similar over the sample size, 2010 to 2014 (Figure 2). This shows that the number of conditional releases in Massachusetts is increasing in relation to the overall prisoner population. However, this means that there has been little if any progress in legislation or policy changes that would lead to more compassionate releases among elderly prisoners in Massachusetts.

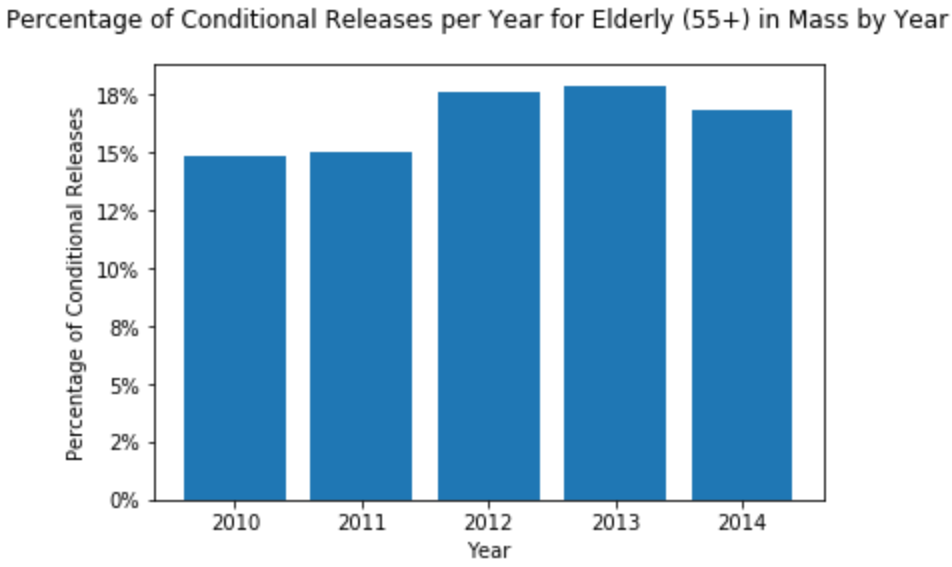

Figure 2. Percentage of Conditional Releases by Year, according to the Prison Release Dataset

In Figure 3, we see the distribution of race among prisoners of different ages by counting individuals in the Year End dataset in Massachusetts, in order to see if the demographics change over time, and if that might cause differences in conditional releases of different races. We see that in Massachusetts, the majority of elderly prisoners are white, which corresponds to the overall population demographics of the state. There is small variation, but because of the small sample size we find that these variations are not significant, and thus the number of individuals of a

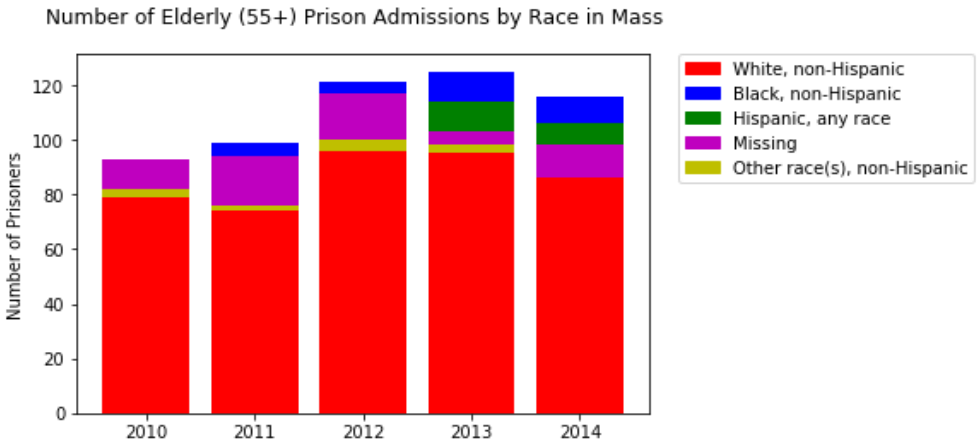

Figure 3. Number of Elderly Prisoners in Massachusetts by Race

certain race does not change over time.

We also looked at the percent of recidivists among prisoners in Massachusetts by looking at the Term Records dataset, as individuals with two entries were indicted twice (Figure 4). From this data we see that elderly prisoners are less likely to enter jail again, and thus are less of a threat to society. This would show that compassionate release programs for elderly prisoners are not as likely to increase crime rates compared to other groups.

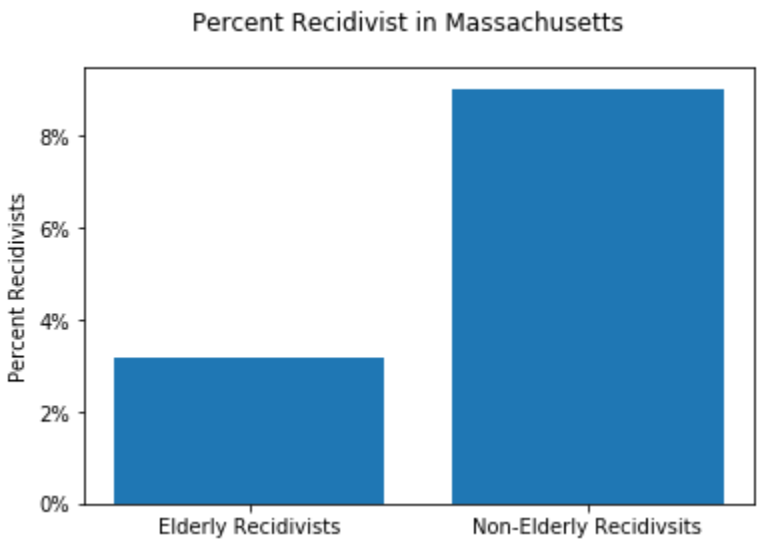

Figure 4. Percent of Elderly Recidivists in the Elderly Prisoner Population compared to Percent of Non-Elderly Recidivists in the Non-Elderly Prisoner Population

Lastly, we found the most common crimes among recidivists that were elderly and not, to see if there were differences in the types of crimes committed between the two groups. As we can see in the table below; non-elderly recidivists were more likely to commit violent crimes (Table 1). It should also be noted that property crimes are composed of burglary, larceny, motor vehicle theft, and arson, which are inherently more violent crimes than public order, which refers to breaking the public peace.

Table 1. Comparing Crimes of Elderly Recidivists and Non-Elderly Recidivists

Percentage of most common crimes among elderly and non-elderly recidivists in Massachusetts	
Elderly recidivists	Non-elderly recidivists
Violent (26%)	Violent (36%)
Drugs (26%)	Drugs (27%)
Public Order (24%)	Property (23%)

Algorithms

In this project we specifically used two main algorithms: a least squares polynomial fit, to predict the number of elderly prisoners in Massachusetts in the future, and a logistic regression to determine the significance of race, level of education, and offense on the likelihood that elderly prisoners get a conditional release. We then use the population size determined by the polynomial fit and coefficients determined by the logistic regression to predict the number of individuals with certain parameters, for example: Hispanics with a high school education that get a conditional release.

We initially graphed the least squares polynomial fit to test the relationship between the overall number of elderly prisoners in the US from 1999 to 2014 and in Massachusetts from 2010 to 2014. A least squares polynomial fit was used as we believed that our data had a linear increase of population over the sample years.

In order to see the effects of how race, education, and offense affect a prisoner's release, we decided to perform a logistic regression. The specific package used was statsmodels.discrete.discrete_model.Logit.fit. Race and education level were selected features as they were the two features that gave demographic information, not counting age. If those features turn out to be significant, we can attribute a demographic to having a lower or higher chance of conditional release. The offense also shapes another demographic, in that it informs prison sentence length, as well as being a measure of how dangerous a prisoner is. Performing a regression with offense would allow us to see biases, and determine if they had an effect on a prisoner getting a conditional release. A logistic regression fits our data well because we have mostly categorical data with a binary dependent variable (An inmate can either have a conditional release or an unconditional release). Specifically the method of regression used was Nelder-Mead, as it avoids many collinearity issues.

In order to bring data from NCRP into a format in which the logistic regression can understand, we first gathered only the relevant data we wanted to consider. We searched our data and only considered inmates who left prison as an elderly inmate (55+) and were released at 2013. We decided on using 2013 instead of 2014 because after we created our polynomial fit to visualize the growing elderly population, we noticed a significant drop in population in 2014, which we believe is due to an early cutoff in the dataset.

We then separated this data randomly into two sets: a training set with 80% of the data and a testing set with the remaining 20%. The purpose of this separation is so we do not overfit our regression to our data; meaning, that we separate some of our data to test the accuracy of the model created from the logistic regression. Finally, we converted the training and test sets into a binary matrix that the logistic regression can understand, with pandas.get_dummies().

We also performed a second logistic regression to inquire into how race and offense affect whether elderly become recidivists. In this regression, we compared elderly recidivists in Massachusetts to elderly non-recidivists in Massachusetts. Logistic regression was used for this because the features were also largely categorical.

Experimental Results

We want to show that the growing elderly population in Massachusetts are not as large of a threat to society when compared to prisoners of other ages. First, we will prove that Massachusetts has a clearly significant growth in elderly inmates. The least squares

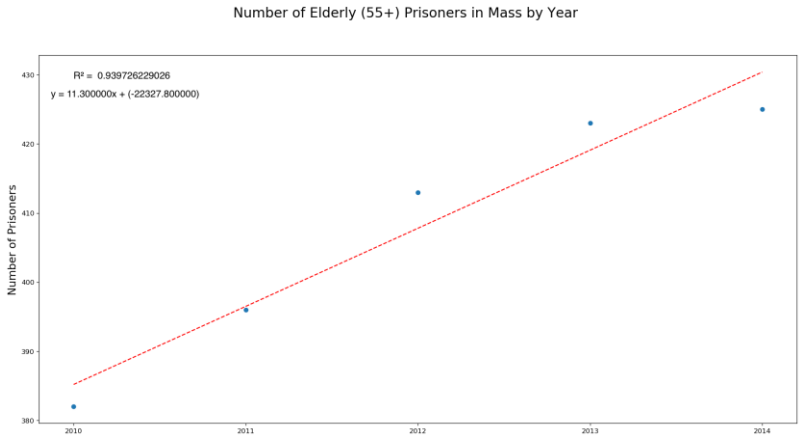

Figure 5. Estimation of the Linear Trend of the Number of Elderly Prisoners in Massachusetts

polynomial fit shows that over time, the overall number of elderly prisoners in Massachusetts increased linearly, with an R^2 value of 0.9397 (Figure 5). We found a similar trend in the polynomial fit of elderly prisoners in the general US (R^2 value of 0.9692) proving that this is a systemic issue in America.

We believe that the elderly should be given more opportunity for conditional release; we explore what factors affect an individual getting a conditional release using a logistic regression (Figure 6). In this regression, the coefficient represents the odds ratio, or the odds of the dependent variable based upon the independent variable; while Z is simply the odds ratio divided by the standard error. The last two columns give the different percentile values of the coefficient. P>|z| category represents the p-value, and we consider a correlation significant if the p-value is less than 0.1.

Dep. Variable:	CONDITIONAL_RELEASE	**No. Observations:**	25198
Model:	Logit	**Df Residuals:**	25176
Method:	MLE	**Df Model:**	21
Date:	Mon, 30 Apr 2018	**Pseudo R-squ.:**	0.01544
Time:	17:58:42	**Log-Likelihood:**	-14484.
converged:	True	**LL-Null:**	-14711.
		LLR p-value:	4.743e-83

	coef	std err	z	P>|z|	[0.025	0.975]
EDUCATION_Under HS diploma/GED	-0.1233	0.039	-3.185	0.001	-0.199	-0.047
EDUCATION_HS diploma/GED	-0.0601	0.036	-1.675	0.094	-0.130	0.010
EDUCATION_Any college	0.3148	0.054	5.856	0.000	0.209	0.420
OFFDETAIL_Murder (including non-negligent manslaughter)	0.6967	0.182	3.830	0.000	0.340	1.053
OFFDETAIL_Negligent manslaughter	0.1592	0.232	0.687	0.492	-0.295	0.613
OFFDETAIL_Rape/sexual assault	0.0860	0.164	0.525	0.599	-0.235	0.407
OFFDETAIL_Robbery	0.3509	0.174	2.012	0.044	0.009	0.693
OFFDETAIL_Aggravated or simple assault	-0.0840	0.165	-0.509	0.611	-0.408	0.240
OFFDETAIL_Other violent offenses	0.2471	0.192	1.287	0.198	-0.129	0.623
OFFDETAIL_Burglary	0.1046	0.166	0.629	0.529	-0.221	0.430
OFFDETAIL_Larceny	-0.5763	0.164	-3.521	0.000	-0.897	-0.255
OFFDETAIL_Motor vehicle theft	-1.1165	0.214	-5.226	0.000	-1.535	-0.698
OFFDETAIL_Fraud	-0.3247	0.173	-1.879	0.060	-0.663	0.014
OFFDETAIL_Other property offenses	-0.1825	0.180	-1.014	0.310	-0.535	0.170
OFFDETAIL_Drugs (includes possession, distribution, trafficking, other)	-0.1199	0.160	-0.748	0.454	-0.434	0.194
OFFDETAIL_Public order	-0.1642	0.160	-1.025	0.305	-0.478	0.150
OFFDETAIL_Other/unspecified	0.0077	0.243	0.032	0.975	-0.469	0.484
RACE_White, non-Hispanic	-0.0383	0.050	-0.766	0.444	-0.136	0.060
RACE_Black, non-Hispanic	-0.1435	0.051	-2.788	0.005	-0.244	-0.043
RACE_Hispanic, any race	0.1672	0.064	2.614	0.009	0.042	0.293
RACE_Other race(s), non-Hispanic	-0.4667	0.090	-5.199	0.000	-0.643	-0.291
Intercept	1.1602	0.163	7.134	0.000	0.841	1.479

Figure 6. Summary of the Logistic Regression with Conditional Release as the Dependent Variable

We found education to be highly significant. If you have a college degree or better you are more likely to be given a conditional release, the reverse is true for high school education or less. This result is intuitive, we would expect educated inmates to better advocate for their release and use prison resources to prove they are no longer a threat. Race is significant to release type for non-white individuals. Being Hispanic lends to a higher likelihood of conditional release, while being other non-white races lends to a lower likelihood of conditional release. This could be the case because Hispanics might be put into jail for longer sentences on average, resulting in more opportunity for parole. Longer sentences have more opportunity to be resolved with conditional releases, as there is more time in prison to demonstrate good behavior. Crimes with a significant effect on release type are murder, robbery, larceny, motor vehicle theft, and fraud. Of these crimes, murder and larceny correlate positively to receiving a conditional release, while the rest do not. This can also be easily explained. Being convicted for larceny, vehicle theft, and fraud do not result in long prison sentences. Average sentence length for larceny is between 2-3 years, with similar sentence lengths for vehicle theft and fraud[1]. There is little to no opportunity to receive parole on a short sentence. Below is a table summarizing these findings (Table 2).

Table 2. Significance of Variables found in the Logistic Regression

Effect on Conditional Release	Significant impact	Non-significant impact
Positive effect	College education or better, Race = Hispanic, Crimes = murder or robbery	Crimes = negligent manslughter, rape, burglary, other violent offenses, other
Negative effect	Less than college education, Race = non-white, non-hispanic Crimes = larceny, vehicle theft, fraud	Race = White Crimes = aggrevated assault, other property offenses, drugs, public order,

We tested this regression by re-performing it with different pairs of the three independent variables. We found that removing even one of the variables had a large change in the significance of the others, indicating that all the variables were very significant to the dependent variable. Having used only 80% of our data to train the regression, we tested the validity of our regression using the remaining 20% of data. We found that our model had a precision of 0.728 and a recall of 0.994. High precision means that our regression returned substantially more relevant results than irrelevant ones, while high recall means that our regression returned most of the relevant results.

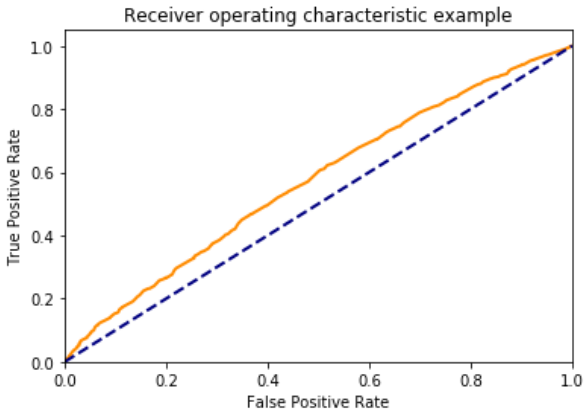

Figure 7. ROC Curve for the First Regression

The Receiver Operating Characteristic (ROC) curve shows how to maximize correct positive calls while trying to minimize false positive calls by manipulating the threshold we use to make these calls (Figure 7). The larger the area under the orange curve, the better our model is. Given the excellent recall and precision score, we would expect our ROC curve to have a larger area, and we are unsure as to why it does not. Overall, we are pleased with the performance of our regression.

Equation 1. Demo Equation Predicting Number of Hispanic Prisoners with a College Degree that get a Conditional Release

$$A(B)\,C \times \frac{1}{1+\exp(D \times 1 + E \times 1 + F)}$$

Using the coefficients determined by the regression and the estimate population size, we can predict the number of people that get released conditionally in a given year based on their race, education, or crime (Equation 1, 2). For

$$498 \times 0.14 \times 0.0035 \times \frac{1}{1+\exp(0.1672 \times 1 + 0.3148 \times 1 + 1.1602)}$$

Equation 2. Filled Variables from Equation 1 with Results of the First Regression

example, let's consider the number of Hispanics in Massachusetts with a college education that receive a conditional release in 2020. Here A is the estimate given by the equation of the line when x=2020, B is the percent of people that are Hispanic, C is the percent of people with a college degree, D is the coefficient given by the logistic regression for Hispanics, E is the coefficient for individuals with college degrees, and F is the coefficient for the intercept. Altogether, this gives us 3.96 Hispanics with college degrees released conditionally in 2020. The creation of this model will be useful for estimating future prison releases.

We now want to take a look at elderly recidivists, and see how race and offense factors into them returning to jail, specifically in Massachusetts, using another logistic regression (Figure 8). We did not consider education because in this small sample the education value was missing for all inmates. Here we found that neither race nor offense had any significance in determining if an elderly inmate would become a recidivist. However, in this regression, precision was only 0.083 and recall was only 0.333. In order to get these readings, we had to use a low threshold of 0.1, meaning that model we developed needs major improvement to become a more accurate predictor. The ROC curve is similar to the previous one, but this does not have much significance considering the inaccuracy of the model.

Dep. Variable:	REPEAT_ID	**No. Observations:**	885
Model:	Logit	**Df Residuals:**	866
Method:	MLE	**Df Model:**	18
Date:	Mon, 30 Apr 2018	**Pseudo R-squ.:**	0.03671
Time:	17:46:54	**Log-Likelihood:**	-203.69
converged:	True	**LL-Null:**	-211.45
		LLR p-value:	0.6258

	coef	std err	z	P>|z|	[0.025	0.975]
OFFDETAIL_Murder (including non-negligent manslaughter)	0.5841	4.749	0.123	0.902	-8.724	9.892
OFFDETAIL_Negligent manslaughter	0.3183	5.830	0.055	0.956	-11.108	11.745
OFFDETAIL_Rape/sexual assault	-0.2344	4.758	-0.049	0.961	-9.560	9.092
OFFDETAIL_Robbery	0.6657	4.760	0.140	0.889	-8.665	9.996
OFFDETAIL_Aggravated or simple assault	0.4018	4.785	0.084	0.933	-8.977	9.780
OFFDETAIL_Other violent offenses	-1.9489	6.977	-0.279	0.780	-15.624	11.726
OFFDETAIL_Burglary	1.5737	4.759	0.331	0.741	-7.753	10.900
OFFDETAIL_Larceny	1.1335	4.771	0.238	0.812	-8.217	10.484
OFFDETAIL_Motor vehicle theft	1.8968	4.858	0.390	0.696	-7.624	11.418
OFFDETAIL_Fraud	-0.9724	5.259	-0.185	0.853	-11.279	9.334
OFFDETAIL_Other property offenses	-0.4446	5.049	-0.088	0.930	-10.341	9.451
OFFDETAIL_Drugs (includes possession, distribution, trafficking, other)	0.4491	4.750	0.095	0.925	-8.860	9.759
OFFDETAIL_Public order	1.2469	4.745	0.263	0.793	-8.054	10.548
OFFDETAIL_Other/unspecified	1.9785	4.857	0.407	0.684	-7.541	11.498
RACE_White, non-Hispanic	-0.3517	0.430	-0.817	0.414	-1.195	0.492
RACE_Black, non-Hispanic	-0.4543	0.521	-0.872	0.383	-1.476	0.567
RACE_Hispanic, any race	-0.2187	0.541	-0.404	0.686	-1.279	0.842
RACE_Other race(s), non-Hispanic	-1.1332	1.950	-0.581	0.561	-4.955	2.688
Intercept	-3.0086	4.723	-0.637	0.524	-12.266	6.249

Figure 8. Results of Logistic Regression with Recidivism as Dependent Variable

Conclusions and Future Steps

We found, definitively, that elderly prisoners were less likely to go back to jail than younger counterparts, and thus they are less likely to cause civil issues after compassionate release programs (Figure 4). We also found that when comparing elderly recidivists commit less violent crimes than non-elderly recidivists, again strengthening our claim that elderly deserve compassionate release programs (Table 1). We were able to first predict the overall population of elderly prisoners in Massachusetts, and use that information in conjunction with a logistic regression to predict the number of individuals with a set of characteristics; in this case, Hispanic with a college degree that get a conditional release in the year 2020. With the conditional release logistic regression we found that being Hispanic, having a college education, and committing violent offenses were very significant in causing an individual to be released conditionally. Committing violent offenses often leads to longer sentences, and thus prisoners have more time to demonstrate good behavior for a conditional release, and if they have some degree of college education, they can better secure conditional release for themselves. However, these effects may be significant because of our

low sample size: not many prisoners in Massachusetts are Hispanic, and not many prisoners get college educations. Our developed model had an excellent recall at 99.4%, and a relatively precise correct positive rate of 72.8%. Our second model, which aimed to determine factors that influence elderly prisoners become recidivists, was less successful with a recall of 33.3% and precision of 8.3%, and needs further development.

For future direction, if we had features for health in our data set, we would love to show that elderly prisoners who are ill are less likely to commit crimes than those who are healthy, and show that there are differences in the crimes that they commit. If those who are sickly commit less serious offenses less often, it would give serious validation to increasing the scale of compassionate release programs in the US. In addition, if we had unique identifiers for prisoners across all our data sets rather than just one, we could see if prisoners that got released conditionally were more or less likely to commit crimes when compared to prisoners that got unconditionally released. Showing this would again show that compassionate release programs would not have a large impact on crime rates across the country.

Citations

Data Source:

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. National Corrections Reporting Program, 1991-2014: Selected Variables. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-09-07. https://doi.org/10.3886/ICPSR36404.v2

1. "Pretrial, Prosecution, and Adjudication." Bureau of Justice Statistics (BJS), www.bjs.gov/content/dcf/ptrpa.cfm.

GitHub

https://github.com/OzdemirCem/CS506FinalProject