A Survival Analysis of Prisoners at the Androscoggin County Jail

Michael Hartnett*

April 2020

Community Engaged Senior Thesis

Submitted to the Department of Economics

Bates College

In Partial Fulfilment of the Requirements for the Degree of Bachelor in Arts

*I would like to thank Bruce Noddin and the Maine Prisoner Re-Entry Network, the Androscoggin County Jail, and my faculty advisor Jonathan Kurzfeld

1 Introduction

1.1 Frequent Users

It is well known that in the criminal justice world there exists a small sub-population of criminals that are known as frequent users.[1] These individuals are regular re-offenders, getting arrested again shortly after every time they are released. When they are not in jail, frequent users also generally struggle to find stability in their daily lives. Anecdotal evidence suggests that frequent users often suffer from mental health and drug abuse disorders. As a result, they rely heavily on resources like homeless shelters and emergency rooms just to survive when they are not in prison. Empirical studies attempting to estimate the size and cost of this population are tremendously difficult primarily due to the difficulty of collecting meaningful data. Those studies that do exist generally focus on identifying these populations locally to demonstrate the severity of the problem (Atkins, Burkhardt, and Lanfear 2016; Torrey et al. 2010).

There is an immense amount of research dedicated to exploring some of the contributing factors of recidivism. The effects of stable housing, schooling, employment, and more been explored by numerous researchers(e.g. Lee 2019; Cook and Kang 2016; Kim, Rabbitt, and Tuttle 2019). However, frequent users often suffer from many of these issues at the same time, and due to administrative problems it can be difficult for the organizations that can provide these different services to coordinate and provide the required help (Stevenson 2017).

In this paper, I use survival analysis techniques to estimate the time between intakes for prisoners at the Androscoggin County jail. By taking advantage of individual level intake data, I am able to estimate the differences between the probabilities of re-offense across different types of crimes. My findings add to the general understanding of recidivism, focusing on individual predictive factors that could also be used to help identify frequent fliers early.

[1] Names for this group of people can vary, they are often referred to as frequent fliers

1.2 Maine

This study uses data from the Androscoggin County Jail (ASJ) in Auburn, Maine. As such, it is important to have an understanding about the circumstances that produced these data. The ASJ is the primary jail for the Auburn PD, the Lewiston PD, and the Androscoggin County sheriff's office. Lewiston and Auburn are the second and fifth largest cities in Maine respectively, and combined have a population of about 55,000 people (*Maine Demographics* 2019). In the two cities, about 18% of people live in poverty, which is higher than the national average of 11.8% (Smega et al. 2018). According to Uniform Crime Reporting data, the crime rates per 1,000 people in Lewiston and Auburn are both close to average compared to national levels, but are higher than other cities in Maine("Crime in Maine" 2018).[2] Like the rest of the country, Lewiston and Auburn have both seen a rise sudden rise in opioid use as well (National Institute on Drug Abuse 2020; Dave, Deza, and Horn 2018).

Neither Lewiston nor Auburn seem to have any particular qualities that would ruin the potential for external validity from the results of an analysis of the ASJ. That is not to say that these findings do carry external validity, just that there is nothing too different. It is perhaps interesting that although these cities have higher than average poverty rates they have succeeded in not having higher than average crime rates, but understanding that relationship fully is beyond the scope of this paper.

[2] Auburn has higher crime rates than Lewiston, but the cities are so close together that it is reasonable to think that the level of crime is not different between them. Auburn just has a smaller population so the crime rate is inflated.

2 Data

2.1 Background

The data I used for this research were individual intakes at the ASJ. The data cover every intake and over three years from 2017 - 2019. Importantly, when an individual is brought to the ASJ for the first time, they are assigned a unique inmate ID. Because of this, I am able to identify when the same person was brought back to the ASJ multiple times over the course of three years. Additionally, when someone is brought to the jail the person responsible for processing their intake includes a brief description of the crime that they were brought in for. It is important to note that the description that is included in the intake records does not necessarily reflect the crime that an individual committed, nor if that individual is actually guilty of having committed any crime in the first place. One of the drawbacks of this data set is that there is not a good way to account for the social costs associated with any of the crimes committed because this method of recording it is potentially inaccurate. However, it is safe to assume that any good police officer would not bring someone to jail unless they had a very good reason to, so we should expect that these descriptions closely represent the harm done to society.

I subdivided the descriptions of crimes into four main categories in order to parse out the different effects on recidivism; violent crimes, drug crimes, property crimes, and other types of crime. The "other" category mostly covers technical violations such as failing to appear for court that often have a comparatively low social cost but still require significant resources from the ASJ and the local police departments. Additionally, I chose to use the "other" category as a way to define certain crimes that did not rise to the severity required by the other categories. For example, driving a car while under the influence of alcohol is extremely dangerous and could result in extreme harm or loss of life, but if someone is pulled over and arrested before this happens, it would hardly be fair to qualify that as a violent crime. Following a similar train of thought, although alcohol is a drug, drunk driving has neither the same motives nor consequences as other drug crimes like distribution. Although the categorization of crimes was subjective, the vast majority were extremely straightforward to decide. There were almost no issues determining whether or not an offense was a drug or property crime. Most of the discretion came when determining whether or not an action rose to the level of being a violent crime. To make these decisions, I referred to the specific section of law that the description mentioned, and took into account whether or not the intake was listed as a felony or misdemeanor.

Another important note about these data is that there is no information about any previous arrests these individuals may have had. There is abundant research that suggests that spending time in jail can negatively affect several outcomes for an individual, thereby increasing the probability of recidivism (Gendreau, Cullen, Goggin, et al. 1999). Unfortunately, this means that not everyone is going to enter the study with the same probability of re-offense. However, although the results might be biased, the result of this will be that individuals with one intake in the sample may have a higher probability of recidivism than individuals who are truly arrested for the first time. In effect, this will cause the estimated results for individuals with fewer intakes in the sample to more closely resemble the results of individuals with more intakes. This will reduce the likelihood that we find significant results, thus helping validate and results we do find.

2.2 Descriptive Statistics

Offense Type	2017	2018	2019	Total
Violent Crime	428	452	395	1275
Drug Crime	281	290	265	836
Property Crime	396	532	449	1377
Other Crime	2514	2324	2351	7189
Total	3619	3598	3460	10677

Table 1: Frequencies of Offense Type

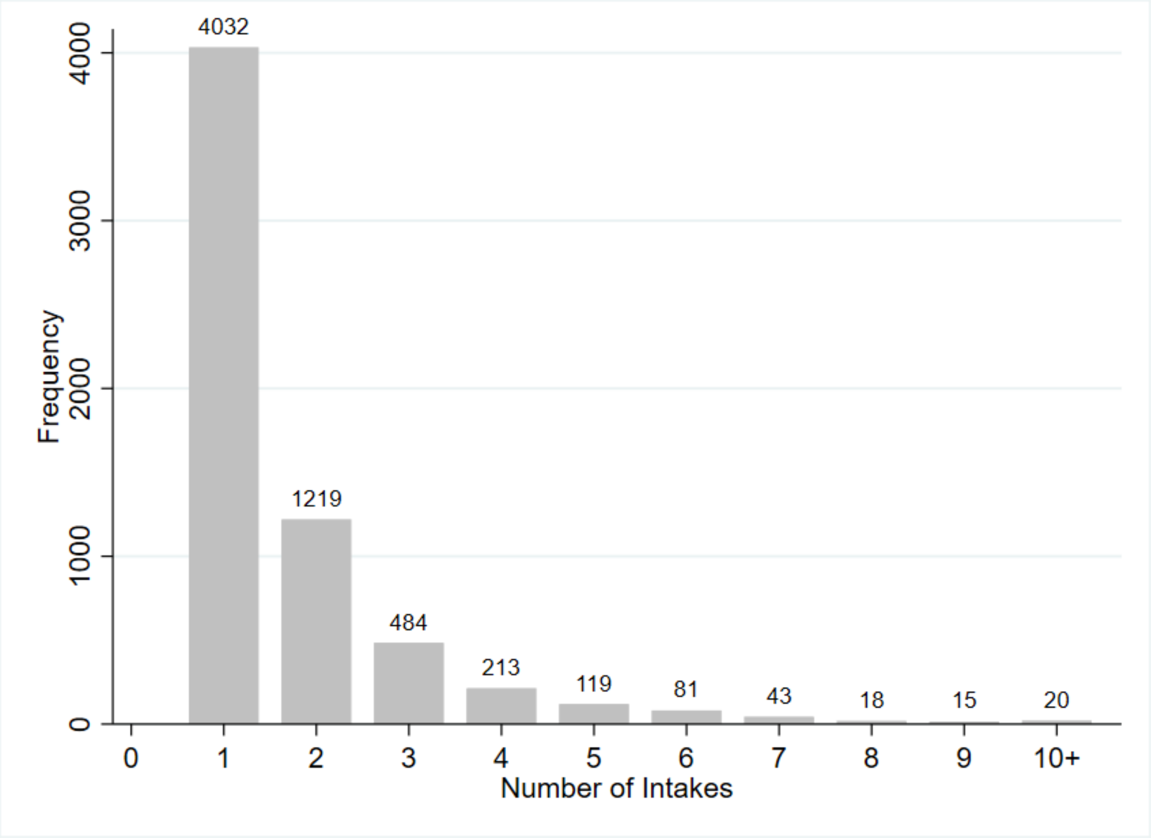


Figure 1: Unique prisoners sorted by number of intakes (N = 6,244)

Table 1 shows the breakdown of intakes by offense type over the three year period. Unsurprisingly, the largest category by far is other crime. Another interesting point is that the number of intakes at the ASJ was fairly constant over this period. There are no particular jumps that might cause concern in our analysis.

We should remember that these categories do not necessarily provide perfect information about the types of crime that happened in the Lewiston-Auburn area over this three year span. These categories only show the reasons that people were brought to the ASJ. With that in mind, these data do provide a good sense of the resources that had to be expended by law enforcement over this time.

Figure 1 breaks down the number of unique intakes that each person in the sample had over the three year period.[3] Over two thirds of the individuals brought to the ASJ did not come back over the window of observation. The unique inmates that averaged more than one intake every year (i.e. at least 4 intakes) only make up about 8% of the population that experienced an intake at the ASJ, but they account for about 23% of the total intakes. This trend seems to suggest that if we could identify these individuals and provide support to reduce their probability of recidivism, then we could cut out a large portion of the intakes.

3 Survival Analysis[4]

3.1 Theory

The goal of survival analysis is to try and understand how certain factors affect the probability that some event of interest happens at a particular time (David Roxbee Cox and Oakes 1984). Generally, we consider a population of people that is at risk of experiencing an event. Mathematically we can say that there is some random variable T that is the time at which the event occurs to someone, and it has some unknown distribution $f(t)$. We say that $f(0)$ is when this person first becomes at risk for this event. Given this, there are two main equations of interest that we want to be able to estimate from our data,

$$h(t) = \lim_{\Delta t \rightarrow\ 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} \tag{1}$$

$$s(t) = P(T > t) \tag{2}$$

[3] The most intakes any individual had in these data was 16.

[4] It is called survival analysis because this type of study is most frequently done in the biomedical field where they might be testing the outcome of some treatment on the probability of death in a patient.

Equation 1 is called the *Hazard Function*, and it gives us the probability that the event of interest occurs at a single moment in time conditional on the fact that the event has not already happened. Equation 2 is called the *Survival Function*, and it shows the probability that the event has not occurred at some given time t.[5] By estimating these two, we not only improve our understanding of how likely an event is to occur, but also better understand how long it should take for that event to occur.

There are a number of different methods that have been proposed for how to estimate these two functions that fall into three main categories: non-parametric, semi-parametric, and parametric. Non-parametric models do not make assumptions about the distribution $f(t)$ and only rely on the observed data. Because these models have the easiest assumptions to meet they are some of the most frequently used, although some critics have found that when it is possible to meet the assumptions of parametric models, they can provide more insight (Miller Jr 1983). The most common semi-parametric model is certainly the Cox proportional hazard model. Although this models still does not assume what the distribution of T is, it assumes that any changes to the hazard function are strictly proportional to each other. We will explore this concept more fully in section 3.4. Finally, parametric models assume the distribution of T and as a result allow for the greatest level of control over the model specification. The obvious draw back is that you must assume your data follow a particular distribution which if inaccurate may invalidate the findings.

In the criminal justice world, survival analysis can be used as a way to better understand recidivism. Instead of simply measuring whether or not a person re-offends, we can use the additional information about how long it takes them to re-offend in order to gain additional insight (DeJong 1997). For these data the event of interest is whether an individual has an additional intake at the ASJ. In order to estimate the hazard and survival functions from these data, we have to make a

[5] $\frac{ds}{dt} \leq 0$ because if t_0 is less than t_1, then $P(T > t_0)$ must be less than $P(T > t_1)$.

few assumptions. First is that an individual first becomes at risk of failure on the day of their first intake at the ASJ in these data. After becoming at risk, an individual remains at risk until either they have a second intake (they fail) or the window of the study expires (they are censored)[6]. If an individual has a second intake, then from that point they are at risk of a third intake (second failure), and this cycle repeats until the end of the study. This method of setting up the model has a few drawbacks. First, it is perhaps unfair to think that an individual is at risk for a second intake starting on the day of their first intake. If someone is arrested and sentenced to a few months in prison, then obviously they are not at risk of being re-admitted to jail during that time. However, because we do not have perfect information about what happens to individuals after their intakes, we are forced to assume that this is when their risk starts. The most important consequence of this assumption is that the hazard rates will be shifted to the right. If we imagine that an individuals is going to be arrested x number of days after they are released, then assuming they are at risk from the day of their initial intake just adds the length of their sentence to x. This is most significant if an individual's sentence length pushes their release back far enough where instead of having another failure in the data, they are censored.

3.2 Log-rank Test

It is especially important with most types of survival analysis to know whether or not the individuals in our data actually have the same hazard function. Most types of analysis rely on the assumption that at each particular point in time, every member of the sample should have the same probability of failure. If this is not true, then we have to adjust our assumptions down the line once we start estimating the hazard function in order to account for these differences.

The log-rank test is one such method of determining whether or not different sub-groups

[6] Censoring is important in survival analysis because although we do not know whether an individual experiences a failure or not, we do know that until that point they had not failed.

	(1)	(2)
Number of Intakes	Events Observed	Events Expected
1	2212	3202.98
2	991	760.66
3	509	252.44
4	296	103.11
5	177	55.03
6	96	27.34
7	53	13.70
8	35	5.451
9	20	5.14
10+	42	5.08
$\chi^2_9 = 276.47$		
$p = 0.000$		

Table 2: Log-rank Test

share the same hazard curve (Mantel 1966; R. Peto and J. Peto 1972). To do this, say our null hypothesis is,

$$h_0 : h_1(t) = h_2(t) = ... = h_{10}(t) \tag{3}$$

The log-rank statistic is very closely related to the χ^2 distribution, which in this case has 9 degrees of freedom. This means we reject our null hypothesis if our χ^2 value is sufficiently large (LaMorte 2016). In practice, the test looks extremely similar to a χ^2 goodness of fit test. Table 7 shows that we can safely reject the null hypothesis that the having different numbers of previous intakes does not change the hazard function for an individual.[7]

3.3 Non-Parametric Estimates

We can begin estimating our hazard and survival functions non-parametrically. The most common non-parametric estimate is the Kaplan-Meier (KM) survival curve (Kaplan and Meier 1958; Clark, Michael J Bradburn, et al. 2003). After adjusting for the number of previous intakes, we can calculate the KM curve by looking at the percentage of the population that fails or is censored on

[7] I also performed several other variations of the log-rank test using different groupings as robustness checks and they all produced similarly strong results.

each day.

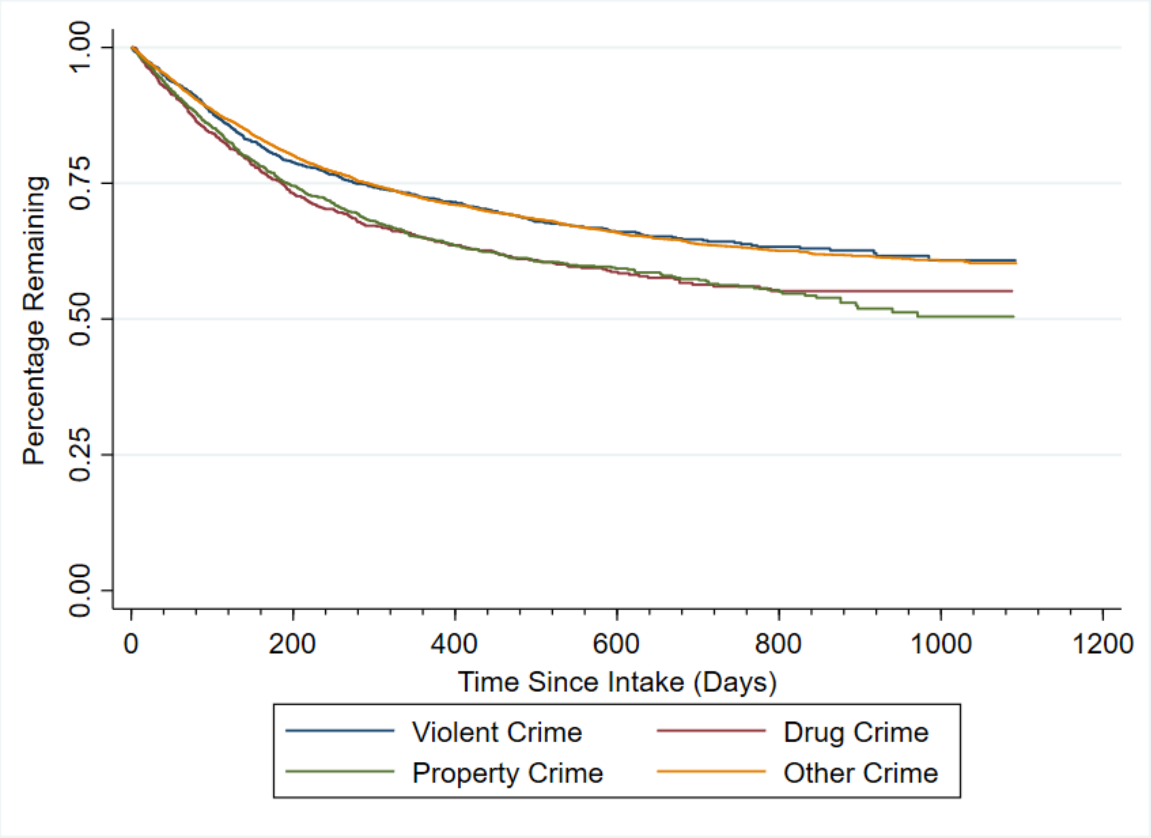
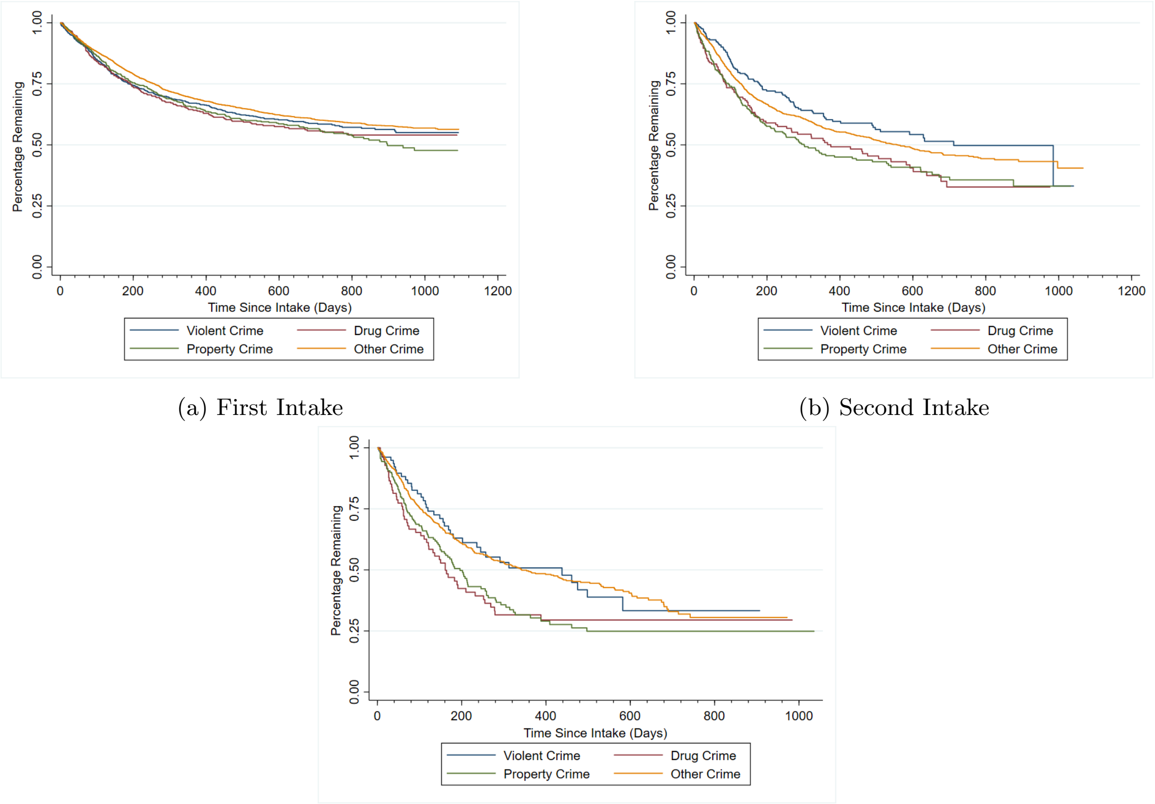


Figure 2: Kaplan-Meier Survival Estimates by Offense Type

Figure 2 shows the adjusted KM curves calculated for for each offense type. We can see that over the whole sample of 10,677 intakes, just over half of the people brought to the ASJ never experienced a second intake. Individuals who committed drug and property crimes were less likely to survive when compared to the other categories for most of the sample window, although the drug crime curve flattens out around 800 days after intake. We can see the estimated KM curves for the first 3 intakes separated out in figure 3. The first intake curve largely matches the whole sample curve, but we can see that as individuals have more intakes, the total survival of the group tends to decrease.



(a) First Intake

(b) Second Intake

(c) Third Intake

Figure 3: Kaplan-Meier Survival Estimates for the First Three Intakes by Offense Type

These three graphs have some interesting implications. By the second and third intakes, we begin to see more clearly a gap between form where for the first two years after intake, people convicted of drug and property crimes are less likely to survive. The variation the at near the end of these graphs is likely less informative, as the negative effects of being incarcerated perhaps lessen and many unobservable things may have happened to the individuals that change their probability of survival. The sample of first intakes seems to go against intuition, suggesting that there may be little difference in the type of crime on the probability of recidivism, with the exception being property crime the dip in the survival of people who committed property crimes near the end of the study window. There are a few possible explanations for this. This may be the result of imperfect

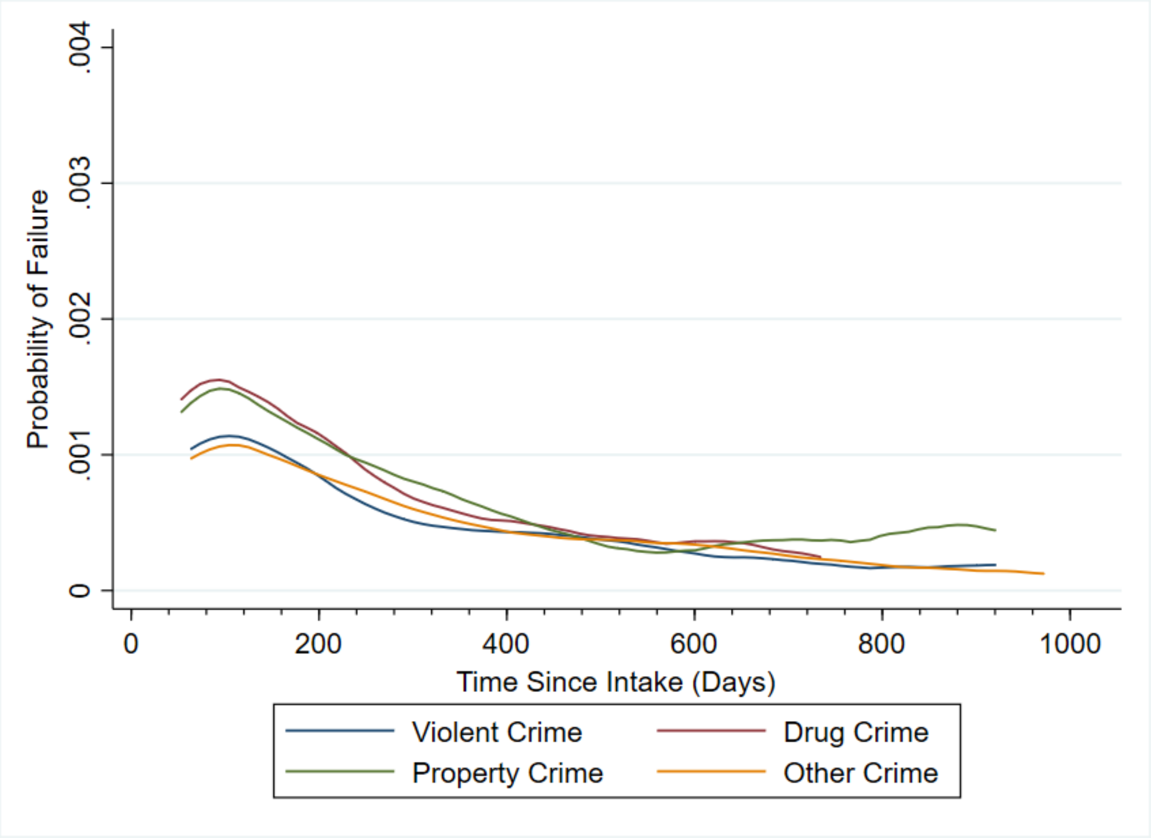
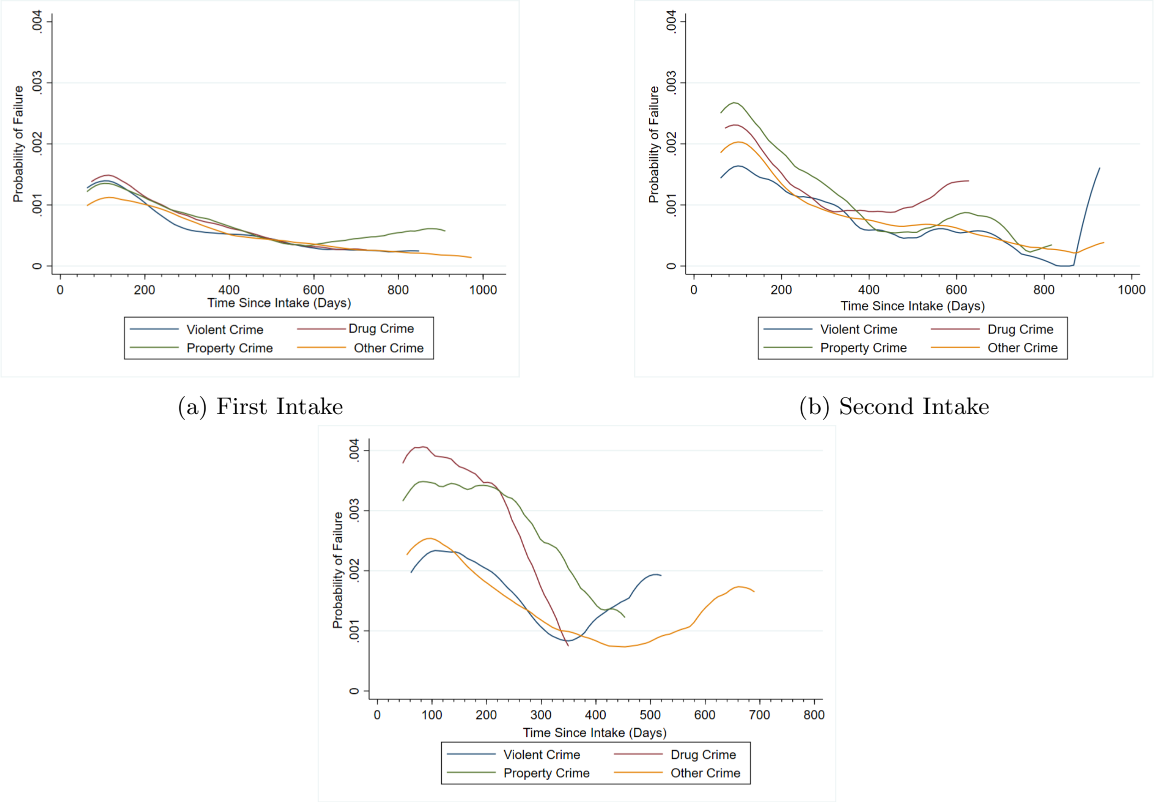


Figure 4: Smoothed Hazard Ratios by Offense Type

information regarding individual's past criminal history that we do not know. If we assume that multiple past intakes increases the probability of re-arrest, then it makes sense that in the first intake of these data we would have the most noise as frequent users with preexisting records are mixed up with people who only get arrested once.

Figure 4 shows the instantaneous hazard rate on a given day since intake. Note that the probabilities on the left hand side are all extremely small. This does not mean that the chance of recidivism is overall very low, but the probability of getting arrested on one specific day is very small. What is more informative is the shape of these graphs. All four of the lines peak at about 100 days since intake. All of this supports the idea that being arrested hurts the most right



(a) First Intake

(b) Second Intake

(c) Third Intake

Figure 5: Smoothed Hazard Functions for the First Three Intakes by Offense Type

after being released. The negative effects on employment, the potential loss of human capital, and strain to social connections among other things makes surviving in the short run significantly more difficult. However, if an individual is able to survive past the first 9 months, then their probability of recidivating decreases. This hypothesis is furthered by the fact that the largest spike is by far for property crimes, which are the most obviously financially motivated. There is a slight uptick in the hazard rate for property crimes near the end of the study window, but again this is less informative because the effects that cause this are likely less due to the previous time incarcerated and may instead be any number of observable things.

When we compare the hazard functions broken down by the first three intakes, we begin to see some interesting trends emerge. The first intake hazard function matches what we saw in the first intake survival curve. All four offense types are grouped quite closely together and follow much of the same trajectory. Again, there is probably a lot more noise in this estimate because we do not have perfect information about the previous criminal histories of each individual. The second intake sample follows much of the same trajectory as the first intake sample, but there is a wider gap between each of the offense types and the hazard rates are noticeably higher at the beginning. The sudden increase in the hazard rates for people who initially committed a violent crime around 900 days after intake is really interesting. It is by far the steepest and jump in any of the figures, and it is unique in that it is the only time the highest hazard rate is not around the 100 day since intake mark. It is hard to find an explanation for this. There do not seem to be any interesting correlations in the data that might cause this spike. The third intake sample shows the clearest picture of the difference between offense types. It makes sense that people who have had multiple previous intakes and were brought in for a drug or property offense are the most likely to recidivate. These are the most likely financially motivated crimes, and after three intakes it is extremely likely that the damage done to an individual's legitimate employment opportunities might be too much to overcome.

This is also perhaps where we can most clearly see the rise of frequent users in the data. The hazard functions for people brought in for drug or property crimes jump dramatically in scale, and they significantly widen the gap between violent and other crimes. It is hard to claim that these people are frequent users, because there is much more than an increased hazard rate that defines that group, but this is a shocking rise. It is perhaps fair to say that if someone experiences a third intake at the ASJ for a drug or property crime, they are at significantly higher risk of becoming a frequent user.

3.4 Empirical Strategy

To estimate the effect of being arrested for a particular type of crime on the probability of being re-arrested at some point in the future, we will use an extended version of the Cox-Proportional Hazards model that allows for repeated events to occur (David R Cox 1972; Andersen and Gill 1982).[8]

$$h(t) = h_{0j}(t) \cdot e^{\beta_1 Violent_i + \beta_2 Drug_i + \beta_3 Property_i} \tag{4}$$

The purpose of this model specification is to try and estimate the proportional effect of the variables of interest on some unknown baseline hazard function h_{0j}, where $1 < j < 16$ denotes the individual baseline hazard functions for subgroups that have the same number of previous intakes in the sample (Clark, Mike J Bradburn, et al. 2003). Because the log-rank test showed that each sub-sample of previous intakes had different hazard functions, we must allow for the baseline hazard to vary across groups in order to satisfy the proportional hazards assumption. Here, the coefficients of interest are on dummy variables that indicate the type of crime individual i committed.[9] These coefficients can be interpreted as the proportion of the indicated group that is likely to have an additional intake on a given day compared to the control. For example, if a people who were brought in for a drug offense have a hazard ratio of 1.50, that would mean that on any given day, people who were initially brought in for drug offenses are 50% more likely to have an additional intake compared to the control group (people who were brought in for other crimes).

We should also note that the timing of financially motivated crime is not random. There is evidence to suggest that crime is related to things like the timing of payments (either income or social benefits) and seasonal changes (Carr and Packham 2019; Hipp et al. 2004). As such, tables 5 and 6 repeat the regression analysis outlined in 3 and additionally include seasonal controls.

[8] Additionally, we use Efron's method for tied failures (Efron 1988)

[9] I choose to exclude the "Other" crimes from the model because the variety of offenses that are included in it make it extremely likely that any coefficient associated with it would not be statistically significant.

Table 3: Regression Outputs from the Cox-Proportional Hazards model

VARIABLES	(1)	(2)	(3)	(4)
Violent Crime	0.153**	-0.184	0.0294	0.0479
	(0.0678)	(0.117)	(0.177)	(0.0511)
Drug Crime	0.301***	0.187	0.204	0.197***
	(0.0839)	(0.130)	(0.179)	(0.0604)
Property Crime	0.268***	0.350***	0.368***	0.270***
	(0.0700)	(0.107)	(0.140)	(0.0486)
Sample	1st Intake	2nd Intake	3rd Intake	Whole Sample
N	6,244	2,210	991	10,673

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Estimated Hazard Rates

VARIABLES	(1)	(2)	(3)	(4)
Violent Crime	1.166**	0.832	1.030	1.049
Drug Crime	1.351***	1.205	1.227	1.217***
Property Crime	1.308***	1.419**	1.446***	1.310***
Sample	1st Intake	2nd Intake	3rd Intake	Whole Sample
N	6,244	2,210	991	10,673

*** p<0.01, ** p<0.05, * p<0.1

4 Results

Table 3 shows the estimated coefficients for the covariates in equation 3. These can be interpreted as the multiplicative effect on the estimated hazard function, which are presented as the hazard ratios in table 4. Tables 5 and 6 show the same results repeated, but with additional controls for the seasonal effects of crime. The hazard ratios can be interpreted as the proportion of the indicated group that we expect to have failed at a given time compared to the proportion of the control group (in this case people arrested for ”other” crimes).

VARIABLES	(1)	(2)	(3)	(4)
Violent Crime	0.122*	-0.189	0.0383	0.0372
	(0.0685)	(0.116)	(0.184)	(0.0513)
Drug Crime	0.266***	0.164	0.200	0.186***
	(0.0846)	(0.132)	(0.180)	(0.0610)
Property Crime	0.230***	0.321***	0.356**	0.261***
	(0.0705)	(0.110)	(0.144)	(0.0491)
Sample	1st Intake	2nd Intake	3rd Intake	Whole Sample
N	6,244	2,210	991	10,673

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Estimated Coefficients with Seasonal Effects

VARIABLES	(1)	(2)	(3)	(4)
Violent Crime	1.129**	0.828	1.039	1.038
Drug Crime	1.304***	1.178	1.221	1.204***
Property Crime	1.259***	1.379***	1.428**	1.299***
Sample	1st Intake	2nd Intake	3rd Intake	Whole Sample
N	6,244	2,210	991	10,673

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Estiated Hazard Rates with Seasonal Effects

Interestingly, when we limit the sample to just the first intakes in in the data, all three coefficients are positive and significant at the 1% level. The more previous intakes that an individual has, the larger the difference becomes for property crimes. It is certainly surprising that after the first intake, individuals who were brought in for drug crimes do not have statistically different hazard ratios. We would certainly expect that because drug crimes are often financially motivated, the more time an individual spends in prison the lower the opportunity cost of committing crime and thus the more likely they are to use those crimes as a source of income. Perhaps it is because financially motivated drug crime has a basic capital requirement. Unless someone is growing or manufacturing their own drugs, they have to be able to buy enough to be able to distribute it. Property crime on the other hand is a much more direct path to financial gain that requires little to no capital upfront. The remarkably large and significant coefficients across all model specifications fits in nicely with the intuition surrounding financially motivated crime.

When we repeat the regressions with seasonal effects included in tables 5 and 6, the coefficients on property crimes begin to decrease slightly. This is to be expected, because we now are able to use seasonal change to measure some of the variation in crime timing. However, the fact that we still have large significant coefficients means that we can be confident in our results.

We can clearly see that with each additional intake, the estimated coefficient for property crimes is increasing. Combining this with the fact that a large percentage of the total intakes in this sample were the result of a small percentage of the population having multiple visits, it appears that by the third intake we may be able to identify the most at risk group. Like all seemingly simple solutions, there is certainly much more we need to understand than what is right in front of us.

5 Discussion

Talking about the factors behind recidivism and finding ways to address it practically can be difficult. It is hard to find efficient ways to identify at risk groups and provide them with meaningful support. It would be foolish to look at these results and think that we could significantly reduce recidivism by taking only the people who on their third time in jail are there for a property crime and doing everything possible to prevent their re-offense. There is no one size fits all solution, just like there is no surefire way to identify people who will re-offend. In addition to the technical challenges of identifying the most at risk groups, we must decide whether or not it is ethical to determine if certain people receive support because they happen to fall into a particular statistical group. If the decision to provide support to a former inmate is simply the result of a statistical analysis, then the more likely it is support goes to the wrong people. The other side to that is of course accurately determining what support each individual needs is near impossible and extremely costly.

Given this, it seems like survival analysis is a tragically underused method when it comes to studying recidivism. The additional information about the timing of re-arrest has the potential to dramatically change policy surrounding providing re-entry support for former inmates. The timing of recidivism is particularly important when it comes to providing services to former inmates. In this paper, the smoothed hazard estimates from figures 4 and 5 regularly showed a large spike around 100 days after the first intake. If these data could be augmented by more information about the timing of release for inmates, then it might be possible to identify a critical window where policy would be the most effective.

Although the specificity of these data allow for interesting research, they also inherently raise questions about the outward validity of these results. Again, neither Lewiston nor Auburn have any particular characteristics that would set them apart from any other small city in the U.S., but it would be wrong to assume that the trends observed here are directly applicable to everywhere else. While we should expect the overall trends to be common across the country, the magnitude of the estimated hazard rates may be subject to change. The best approach to test the validity of these results would be for other researchers to collect data from their local prisons and run similar survival analyses. Additionally, this type of research would certainly benefit from more individual level information. On of the biggest drawbacks of this paper is that because it is an undergraduate thesis, I did not have enough time to collect meaningful supplemental data. The greatest obstacle to this is data access. Perhaps if this paper were to be replicated on a much larger scale, then there could be some sort of weight given to individuals from different prisons that tries account for variations

Number of Intakes	Number of Subjects
1	4032
2	1219
3	484
4	213
5	119
6	81
7	43
8	18
9	15
10+	20
$\chi^2_9 = 276.47$	
$p = 0.000$	

Table 7: Log-rank Test

in mental health. It will be challenging to find meaningful ways to further our understanding of frequent users, but there is so much to learn and so much efficiency to be gained that it certainly a worthwhile pursuit.

6 Conclusion

Because the top percentiles of offenders account for the highest percentage of the intakes, it is important to understand what makes people more likely to fall into that cycle of repeated re-arrest. The method outlined in this paper is a good starting point from which we can grow our understanding of the early warning signs of recidivism and begin to provide more support to prevent re-arrest. Hopefully, continued survival analysis will help identify the most at risk groups across the country.

References

[1] Per Kragh Andersen and Richard D Gill. "Cox's regression model for counting processes: a large sample study". In: *The annals of statistics* (1982), pp. 1100–1120.

[2] Scott Atkins, Brett Burkhardt, and Charles Lanfear. "Law Enforcement Response to "Frequent Fliers: An Examination of High-Frequency Contacts Between Police and Justice-Involved Persons With Mental Illness". In: *Criminal Justice Policy Review* 322.1 (2016), pp. 97–114. DOI: https://doi.org/10.1177/0887403414559268.

[3] Jillian B Carr and Analisa Packham. "SNAP Schedules and Domestic Violence". In: (2019).

[4] Taane G Clark, Michael J Bradburn, et al. "Survival analysis part I: basic concepts and first analyses". In: *British journal of cancer* 89.2 (2003), pp. 232–238.

[5] Taane G Clark, Mike J Bradburn, et al. "Survival analysis part II: multivariate data analysis–an introduction to concepts and methods". In: *British journal of cancer* 89.3 (2003), pp. 431–436.

[6] Philip J Cook and Songman Kang. "Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation". In: *American Economic Journal: Applied Economics* 8.1 (2016), pp. 33–57.

[7] David R Cox. "Regression models and life-tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (1972), pp. 187–202.

[8] David Roxbee Cox and David Oakes. *Analysis of survival data.* Vol. 21. CRC Press, 1984.

[9] "Crime in Maine". In: (2018).

[10] Dhaval Dave, Monica Deza, and Brady P Horn. *Prescription drug monitoring programs, opioid abuse, and crime.* Tech. rep. National Bureau of Economic Research, 2018.

[11] Christina DeJong. “Survival analysis and specific deterrence: Integrating theoretical and empirical models of recidivism”. In: *Criminology* 35.4 (1997), pp. 561–576.

[12] Bradley Efron. “Logistic regression, survival analysis, and the Kaplan-Meier curve”. In: *Journal of the American statistical Association* 83.402 (1988), pp. 414–425.

[13] Paul Gendreau, Francis T Cullen, Claire Goggin, et al. *The effects of prison sentences on recidivism.* Solicitor General Canada Ottawa, Ontario, 1999.

[14] John R Hipp et al. “Crimes of opportunity or crimes of emotion? Testing two explanations of seasonal change in crime”. In: *Social Forces* 82.4 (2004), pp. 1333–1372.

[15] Edward L Kaplan and Paul Meier. “Nonparametric estimation from incomplete observations”. In: *Journal of the American statistical association* 53.282 (1958), pp. 457–481.

[16] Jiyoon June Kim, Matthew P Rabbitt, and Charlotte Tuttle. “Changes in Low-Income Households’ Spending and Time Use Patterns in Response to the 2013 Sunset of the ARRA-SNAP Benefit”. In: *Applied Economic Perspectives and Policy* (2019).

[17] Wayne LaMorte. *Comparing Survival Curves.* 2016. URL: `http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Survival/BS704_Survival5.html`.

[18] Logan M Lee. “Halfway Home? Residential Housing and Reincarceration”. In: (2019).

[19] *Maine Demographics.* 2019. URL: `https://www.maine-demographics.com/cities_by_population`.

[20] Nathan Mantel. “Evaluation of survival data and two new rank order statistics arising in its consideration”. In: *Cancer Chemother. Rep.* 50 (1966), pp. 163–170.

[21] Rupert G Miller Jr. “What price kaplan-meier?” In: *Biometrics* (1983), pp. 1077–1081.

[22] Richard Peto and Julian Peto. “Asymptotically efficient rank invariant test procedures”. In: *Journal of the Royal Statistical Society: Series A (General)* 135.2 (1972), pp. 185–198.

[23] Jessica Smega et al. *Income and Poverty in the United States: 2018.* 2018. URL: `https://www.census.gov/library/publications/2019/demo/p60-266.html`.

[24] Katherine G Stevenson. "Intersections of Mental Illness and Legislative Changes at Androscoggin County Jail". In: (2017).

[25] E Fuller Torrey et al. "More mentally ill persons are in jails and prisons than hospitals: A survey of the states". In: *Arlington, VA: Treatment Advocacy Center* (2010), pp. 1–18.