3-year Recidivism Rates in Parole and Non-Parole Samples of Iowa Prisoners

Evan Moir

22/03/2020

Abstract

Using parole as a treatment compared to a non-paroled control group, 3-year recidivism rates were studied in a set of Iowa prisoners released from 2010 to 2012. Using a chi-squared test, it was shown that there is a significant increase in recidivism in the parole group over the non-parole group. While there are challenges to the internal and external validity of this analysis, it should indicate to justice and prison systems that further study may be needed to understand the relationship between parole and recidivism.

Introduction

The question of how to reduce recidivism in prisoners is of significant interest to those who study justice and prison systems: reducing recidivism means reducing both the costs associated with prison systems, and the social impact on the victims of crimes, the families of prisoners, and prisoners themselves. This analysis approaches the question of which factors affect recidivism by investigating the role of parole - are paroled prisoners more or less likely to be recidivists?

The Dataset

The chosen dataset gives 3-year recidivism for 26,020 prisoners in Iowa released during their prison system's 2010-2015 fiscal years. 3-year recidivism means, e.g., if the prisoner was released in 2011, a follow-up was done in 2014 to see if they'd been readmitted to prison at any point since release. The dataset is provided by the Government of Iowa and was originally found on Kaggle - links to both webpages are in appendix A.

The dataset contains the following variables of interest. Most were renamed from their original names for convenience.

- A boolean capturing whether the prisoner was re-admitted to prison within three years after their 2010 release (recidivist)
- The manner of their release: Parole, Special Sentence, Discharged at End of Sentence (release_type)
- The year of release (release_year)
- The prisoner's ethnicity (ethnicity)
- The prisoner's age range at release (age)
- The prisoner's sex (sex)
- The type of the original convicted offense (offense_type)

Visualizations of the distributions of these key variables of interest can be found in the *Internal Validity* section below.

Research Question

The goal of this analysis is to determine if the manner of a prioner's release has an effect on their probability of recidivism. There are three types of release (Parole, Special Sentence, Discharged at End of Sentence), so there are several possible ways to set up an A/B test. Three dummy variables were created to determine how to best split the sample:

- parole (release_type == "Parole")
- full_sentence (release_type == "Discharged End of Sentence")
- special_sentence (release_type == "Special Sentence")

The size of the samples for these splits will determine which are suitable for an A/B test:

The parole / no parole split is the only one with comparable sample sizes between the two groups. So the broad research question becomes: "What is the effect of parole being granted on 3-year recidivism rates?" Or, more precisely:

Null hypothesis: there is no significant difference in 3-year recidivism rates between prisoners granted parole and those not granted parole.

Hypothesis: there is a statistically significant difference in 3-year recidivism rates between prisoners granted parole and those not granted parole.

In this scenario parole is being used as a treatment, with non-paroled prisoners as the control.

Preprocessing

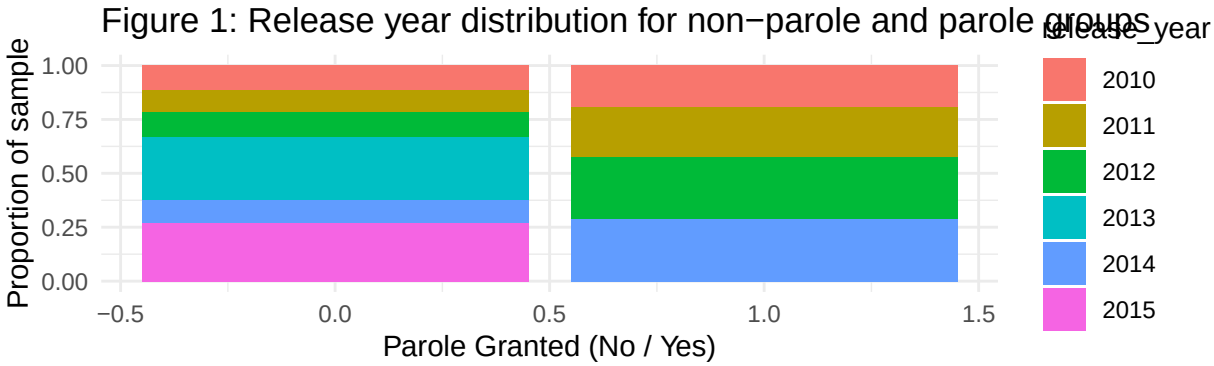
To ensure clean data, blank values were transformed into NAs, and rows with NAs in the following variables of interest were removed: recidivist, age, ethnicity, sex, offense_type, and year_of_release. No NAs were found in these variables.

However, the ethnicity variable contained some ambiguous values, e.g. "White - Hispanic" and "White - Non-Hispanic", but also "White -". To counter this, observations with ambiguous values (e.g."White -") were removed from the dataset. This brought the dataset down to 25,971 rows.

Internal Validity

To attempt to establish internal validity, it's important to show that the two groups are similar along the other dimensions of interest in the dataset. This can be done visually, by comparing the parole and non-parole groups side by side.

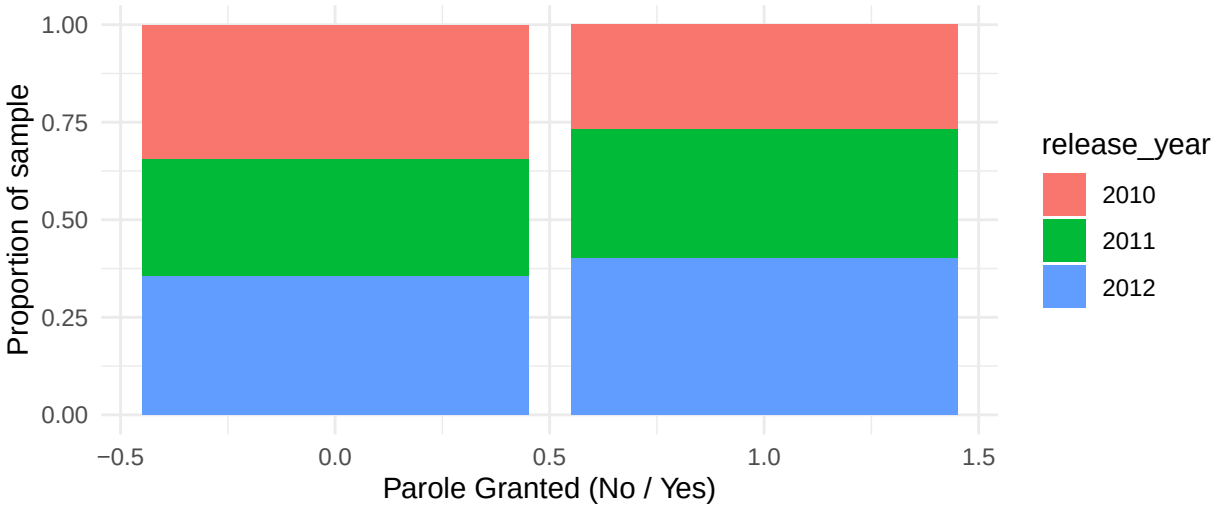
Release Year: It would be ideal to have identical distributions for release year between the two groups, as there may have been policy changes regarding parole, or even new laws that made recidivism more or less likely - but these are unknown to us without significant research and resources.



There is cause for concern here: the two groups only share the years 2010-2012. By keeping only the years shared by both groups (2010-2012), this concern can be sidestepped at the cost of a smaller *but still acceptably large) sample. This leaves 12,327 rows in total, with the following split:* Non-parole: 5,387 * Parole: 6,986

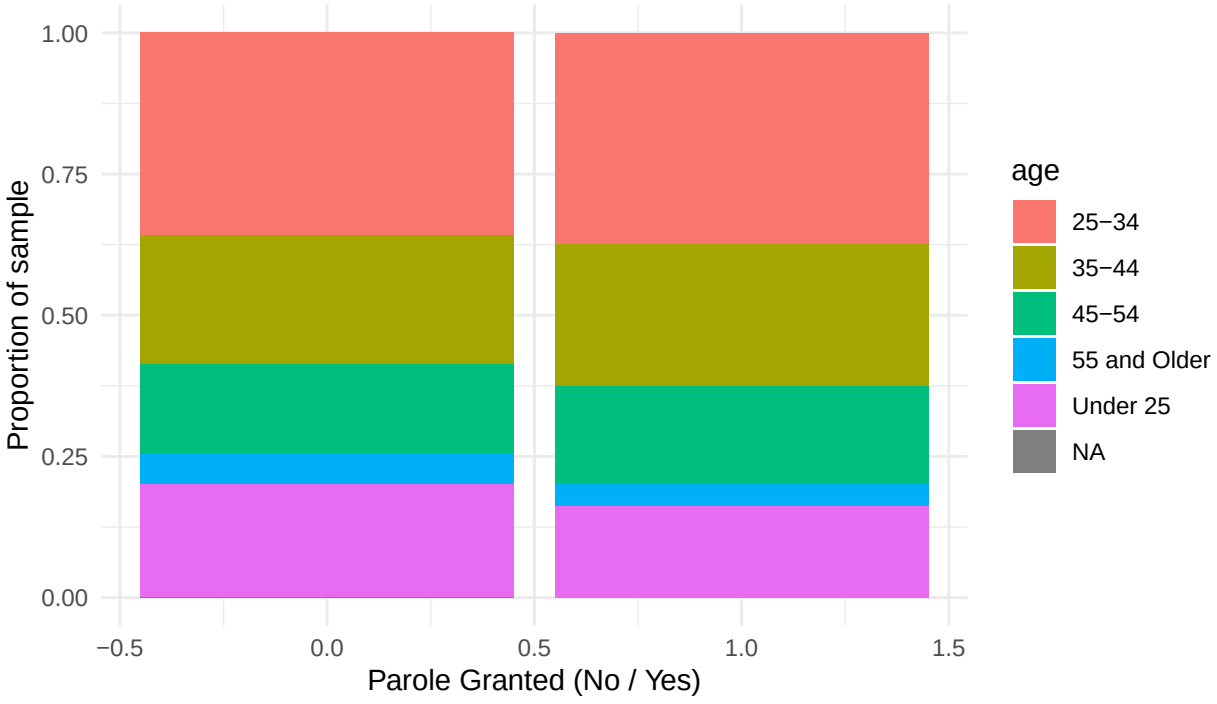
The resulting distribution in release_year is much more acceptable.

Figure 2: Final release year distribution for non−parole and parole groups



Age: Similar distributions here are important here, as there may be age-related bias inherent in the justice/prison system.

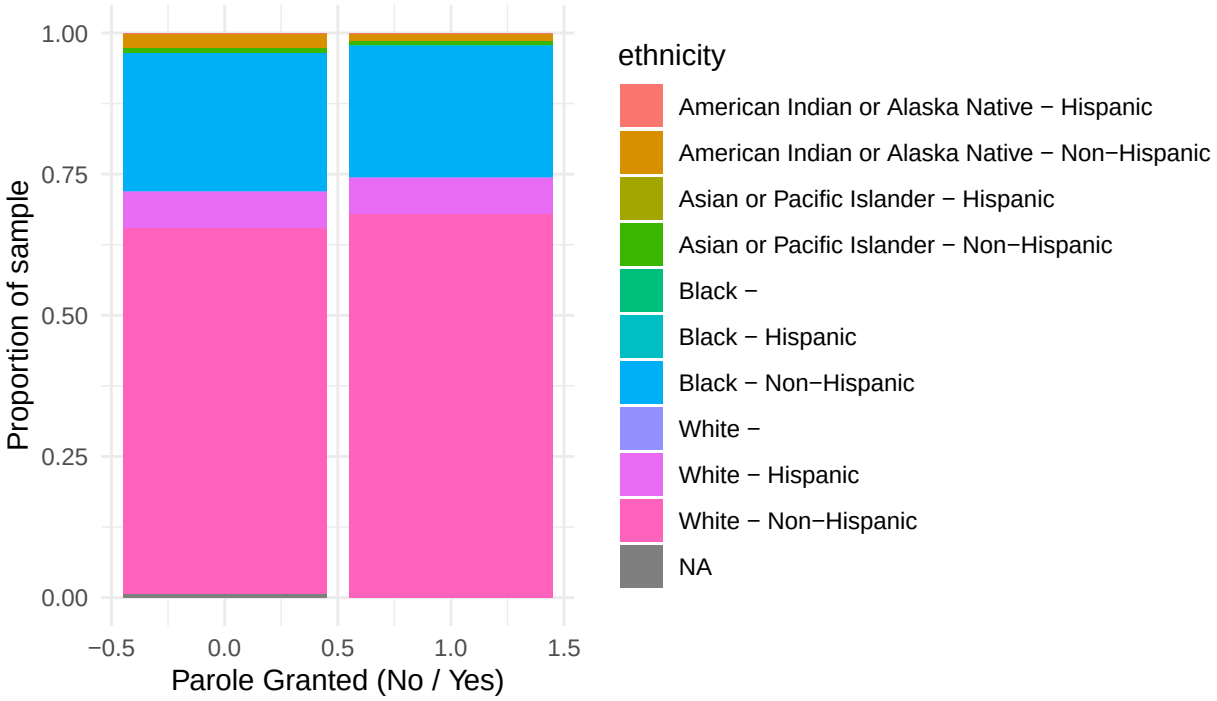
Figure 3: Age distribution for non−parole and parole groups



These distributions are very close, with small but acceptable differences the *Under 25* and *35-44* categories.

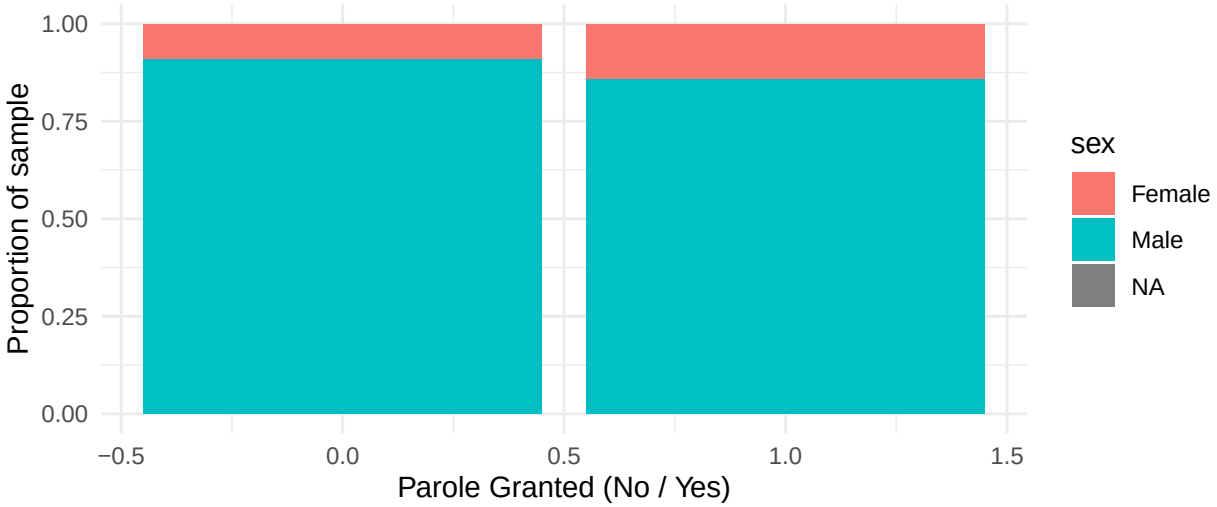
Ethnicity: Similar distributions here are important, as there may be ethnicity-centred biases inherent in the justice/prison system and the decision-making of the individuals therein.

Figure 4: Ethnicity distribution for non−parole and parole groups



These distributions are also quite close, with small but acceptable differences mainly in the *White - Non-Hispanic* and *Black - Non-Hispanic* categories.

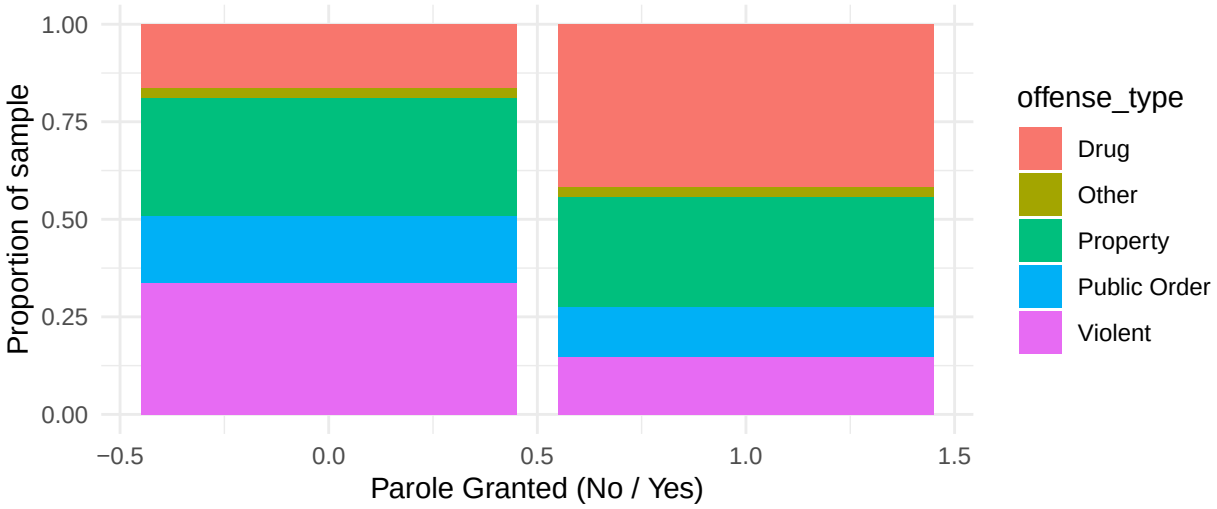
Figure 5: Sex distribution for non−parole and parole groups



Here we see a small but acceptable difference between the sex distributions.

Offense Type:

Figure 6: Offense type distribution for non−parole and parole groups



There is cause for concern here. There are significant differences in the proportion of drug and violent crime offences in the two groups. This is the largest threat to internal validity. The impact of this on the analysis will be elaborated on in the *Discussion* section.

Analysis

Recidivism rate is calculated by:

$$R = n_{recidivist}/n_{sample}$$

Rates for the two groups are summarized in Table 1.

Table 1: Recidivism rates for non-parole and parole groups

Non-parole	Parole	Delta
0.263	0.341	0.078

This suggests that paroled prisoners have a higher recidivism rate than non-paroled prisoners. But is this difference significant? To answer this, a chi-squared test is performed using the parole variable as the treatment and recidivism as the outcome variables. The chi-squared test compares the observed frequencies of recidivism in the two groups to their expected frequencies to evaluate how likely the observed frequencies would be assuming the null hypothesis is true.

The frequencies used in calculating the chi-squared test are shown in table 2

Table 2: Frequency matrix for chi-squared test

	Non-Parole	Parole	Total
Non-recidivist	3968	4606	8574
Recidivist	1419	2380	3799
Total	5387	6986	12373

The chi-squared test resulted in:

- chi-squared: 84.99
- p value < 2.2 * 10^-22.

Using the standard practice of considering p = 0.05 the cut off for significance, the null hypothesis is rejected: there is a statistically significant increase in 3-year recidivism rates in the parole group over the non-parole group.

Recidivism rates for the levels of each of the variables of interest was also calculated for intere's sake, though these subgroups likely have significant internal validation issues. Appendix B summarizes these findings.

Discussion

The analysis led to the null hypothesis being rejected, i.e. a statistically significant increase in 3-year recidivism rates in the parole group over the non-parole group was found. This suggests that parolees are at higher risk for recidivism compared to prisoners serving their full sentence. The reasons for this are not well understood, but it may be because parole sends the signal that full sentences are typically not enforced, making it worth the risk to commit additional crimes.

That said, there was one significant concern regarding internal validity: the distribution of offense types of the two groups were not comparable. This is a weakness of the analysis, as the type of offense can affect both the probability of parole and the probability of recidivism (and possibly other aspects in the experiment context as well).

It should also be noted that the data does not include any information on the criteria used in parole decisions. Because of this, there are unobserved variables at play.

Given that this data represents only a few years' worth of prisoners from a single prison system, it's hard to argue that this analysis has strong external validity. There are many possible factors that affect the probability of a prisoner re-committing crime, some of which are related to the justice/prison system itself, and others to the society at large. For example, prisoners in a system more focused on rehabilitation, or without the proft motive found in the US prison system, may have different results.

There are also a number factors to consider regarding the personal circumstances of the parolee, e.g.: do they have employment and a social support network? These factors have the potential to play a large role in determining individual outcomes, but were not considered in this analysis.

A follow-up analysis seeking to improve internal and external validity could focus on the following changes:

- Restrict the number of offense types under consideration, e.g. exclude violent crimes.
- Investigate jurisdiction-specific factors relating to how parole is granted.
- Measure aspects of parolees' personal employment situation and support network, e.g. what type of work and housing are available to them during parole.
- Include data on the considerations used in making parole decisions.

Conclusion

While it was shown that there was a statistically significant increase in recidivism among paroled prisoners for this dataset, there were concerns about internal validity, and it's difficult to generalize this result to other justice/prison systems due to the number of factors that can change between such systems. However, the result suggests that parolees may be at higher risk of recidivism than prisoners serving their full sentence of receiving special release conditions. More study is needed.

Ethical Considerations

There are significant ethical considerations anytime a prison system is discussed - prisoners are a vulnerable population at the mercy of the state, justice system and prison system they inhabit, and particularly justice systems are biased in ways that can't always be known without significant study. Particularly in the United States, where prison can be a for-profit venture, there are (arguably perverse) incentives that do not work in favour of the well-being or rehabilitation of prisoners.

Regarding the data itself

The definition of the ethnicity categories is of significant concern. Ethnicity is a slippery concept that is not well defined biologically or in any socially universal sense - it tends to be vary significantly depending on the cultural context in which it's being discussed. For example: why are all of the categories appended with "hispanic" or "non-hispanic"? This speaks to a specifically American viewpoint of ethnicity.

Particularly given that backdrop, there are ethical considerations to even attempting to collect ethnicity as part of a dataset. It has the potential to bias anyone making decisions using the data. IT's intended use in this data is not known.

The same could be true of sex, e.g. male and female prisoners could be treated differently because of unknown biases on the part of the decision-makers. In addition, this dataset has a very binary view of sex. Even if sex is intended in the biological sense (rather than gender), the modern scientific view of sex is not a strict binary. It's unclear what there is to gain from including this as a variable.

Presumably (though not known for sure) prisoners in the dataset are not consenting to their inclusion in this data. In most discussions of ethical data collection, consent plays a big part. While there may be positive intentions behind the collection of the data, and it maybe important for helping future prisoners, the fact that it's unlikely prisoners gave consent remains a concern.

Regarding experimental design

Had an experiment using parole as treatment been carried out prior as part of the collection of the data, there would be significant ethical concerns regarding the fairness of making parole decisions for the purposes of the experiment rather than based on the behaviour and potential threat to society posed by the individuals themselves. Because this was a post-hoc analysis, those potential issues are sidestepped. But this speaks to one of the challenges of this type of study - balancing good and useful experiment design with the ethical considerations and treatment of individuals is a challenge all study designers need to wrestle with.

References

- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- David Robinson and Alex Hayes (2019). broom: Convert Statistical Analysis Objects into Tidy Tibbles. R package version 0.5.3. https://CRAN.R-project.org/package=broom
- Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 0.8.4.https://CRAN.R-project.org/package=dplyr
- Yihui Xie (2019). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.26.

Appendix A

The Dataset

Courtesy of the Government of Iowa:

3-Year Recidivism for Offenders Released from Prison in Iowa

Kaggle page where the dataset was originally found:

Recidivism for Offenders Released from Prison

Full code for reproducibility

Click here for link to GitHub repository

Appendix B

Table 3: Recidivism rates for non-parole / parole groups by sex

Sex	No Parole	Parole	Delta
Male	0.273	0.348	0.075
Female	0.016	0.293	0.277

Table 4: Recidivism rates for non-parole / parole groups by release year

Release Year	No Parole Group	Parole Group	Delta
2010	0.245	0.361	0.116
2011	0.272	0.314	0.042
2012	0.274	0.349	0.075

Table 5: Recidivism rates in non-parole / parole groups by Offense Type

Offense Type	No Parole Group	Parole Group	Delta
Drug	0.248	0.355	0.107
Property	0.282	0.388	0.106
Public Order	0.246	0.285	0.039
Violent	0.252	0.249	-0.003
Other	0.065	0.023	-0.042

Table 6: Recidivism rates for non-parole / parole groups by age

Age Range	Non-parole	Parole	Delta
Under 25	0.322	0.385	0.063
25-34	0.271	0.363	0.092
35-44	0.249	0.330	0.081
44-54	0.223	0.291	0.068
55 and Older	0.176	0.239	0.063

Table 7: Recidivism rates for non-parole / parole groups by ethnicity

Ethnicity	Non-parole	Parole	Delta
American Indian of Alaska Native - Hispanic	0.000	0.500	0.500
American Indian of Alaska Native - Non-Hispanic	0.000	0.473	0.473
Asian or Pacific Islander - Hispanic	Inf	NaN	NaN
Asian or Pacific Islander - Non-Hispanic	0.292	0.229	-0.063
Black - Hispanic	0.200	0.300	0.100
Black - Non-Hispanic	0.214	0.356	0.142
White - Hispanic	0.136	0.206	0.070

Ethnicity	Non-parole	Parole	Delta
White - Non-Hispanic	0.202	0.347	0.145