Application of Survival Analysis

This analysis draws on data from an experimental study on recidivism, involving 432 male individuals who were tracked for one year following their release from prison. The dataset includes a broad set of demographic, socioeconomic, and criminal history variables to evaluate their potential influence on the time to recidivism.

<u>Variables included in the analysis</u>

- *week*: Number of weeks from release until either the first re-arrest or the end of the observation period.
- *arrest*: Indicator variable for whether the individual was arrested during the study period (1 = yes)
- *fin*: Indicator for receipt of financial assistance post-release (1 = yes). Financial aid was randomly assigned by the researchers.
- *race*: Indicator for race (1 = Black).
- *wexp*: Indicator for full-time work experience prior to incarceration (1 = yes).
- *mar*: Indicator for marital status at release (1 = married).
- *paro*: Indicator for parole status (1 = released under parole).
- *prio*: Number of prior convictions.
- *educ*: Education level, categorized as:

    - 2 = Grade 6 or less
    - 3 = Grades 6–9
    - 4 = Grades 10–11
    - 5 = Grade 12
    - 6 = Some post-secondary education

The main objective of this study is to identify factors associated with the timing and likelihood of recidivism using survival analysis techniques, including Kaplan-Meier estimates and Cox proportional hazards models.

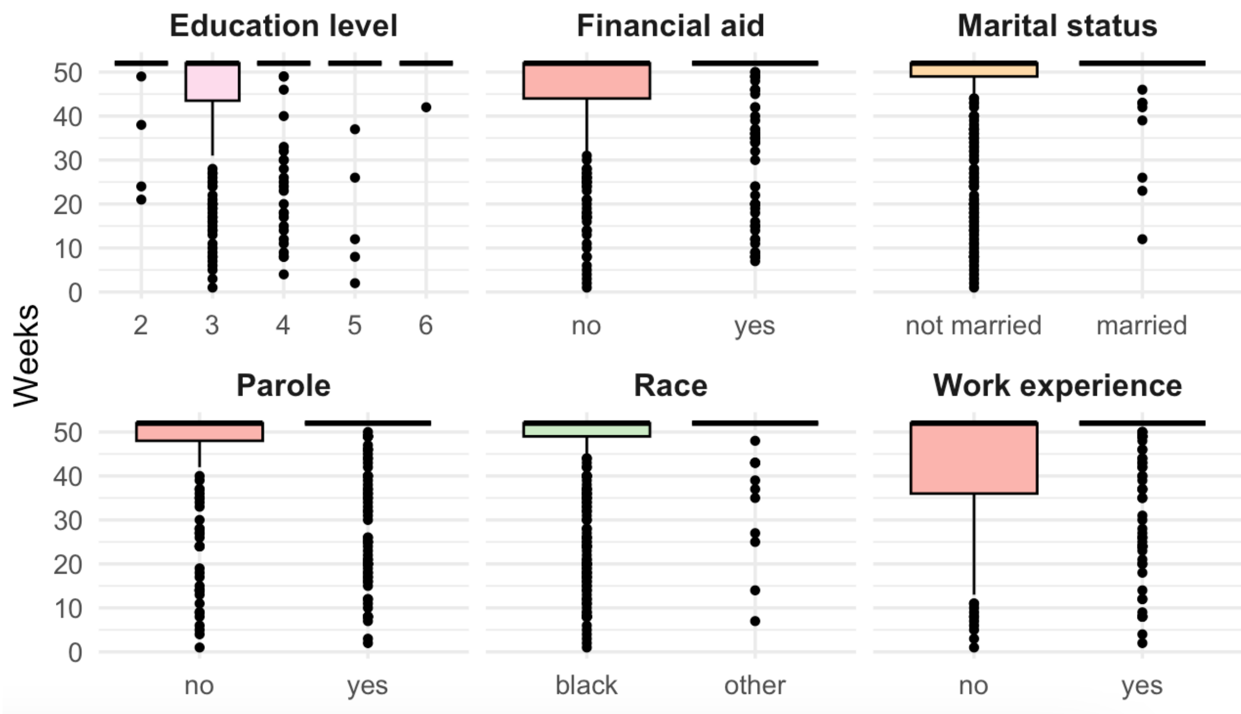
The report begins with an exploratory analysis of the time to recidivism, disaggregated by key individual characteristics such as education, marital status, financial assistance, work experience, race, and parole status. This section uses summary statistics, boxplots, and smoothed scatterplots to highlight initial patterns and potential disparities across groups. It then moves into formal statistical testing using log-rank tests and Kaplan-Meier survival estimates, followed by multivariate modeling through the Cox proportional hazards framework. Together, these methods provide a comprehensive view of how different factors shape the likelihood and timing of reoffending and help identify potential entry points for targeted interventions.

Exploratory Analysis of Time to Recidivism

This section presents an exploratory overview of the distribution of time to recidivism based on key individual characteristics. Categorical variables were assessed using boxplots, while continuous variables were examined through smoothed scatterplots. This mixed-methods approach helps uncover initial trends and potential associations between predictors and the timing of reoffending.

Preliminary results are illustrated in Figure 1, which displays the distribution of time to recidivism (in weeks) by education level, marital status, financial assistance, work experience, race, and parole status.

Figure 1. Time to Recidivism by Individual Characteristics



Complementing the visual analysis, Table 1 provides summary statistics, including sample size, minimum, quartiles, maximum, mean, and standard deviation for each subgroup. The median is consistently 52 weeks across all groups, reflecting that a large share of individuals did not recidivate within the one-year observation period. However, differences in the lower quartiles and overall dispersion reveal important variation in the timing of recidivism, particularly among individuals with different socioeconomic backgrounds. The following paragraphs examine these patterns in more detail, beginning with differences observed across educational attainment.

Table 1. Summary Statistics – Time to Recidivism by Individual Characteristics

Characteristic	Group	N	Min	Q1	Median	Q3	Max	Mean	SD
Financial Assistance	No	216	1	44	52	52	52	44.8	13.5
	Yes	216	7	52	52	52	52	46.9	11.7
Race	Black	379	1	49	52	52	52	45.6	13
	Other	53	7	52	52	52	52	48	9.73
Work Experience	No	185	1	36	52	52	52	43.2	14.6
	Yes	247	2	52	52	52	52	47.8	10.6
Marital Status	Not married	379	1	49	52	52	52	45.4	13.1
	Married	53	12	52	52	52	52	49.3	7.85
Parole Status	No	165	1	48	52	52	52	45.4	13.2
	Yes	267	2	52	52	52	52	46.1	12.3
	2 (≤ Grade 6)	24	21	52	52	52	52	48.8	8.62

Education Level	3 (Grades 6–9)	239	1	43.5	52	52	52	45	13
	4 (Grades 10–11)	119	4	52	52	52	52	46	13
	5 (Grade 12)	39	2	52	52	52	52	47.5	12.7
	6 (Post-secondary)	11	42	52	52	52	52	51.1	3.02

Educational attainment exhibits noticeable distinctions across groups. Although the median is 52 weeks in all categories—indicating that most individuals remained offense-free throughout the observation period—differences emerge when looking at the lower quartile and the variability in time to recidivism. Group 3 (Grades 6–9), which accounts for the largest subgroup (n = 239), shows a lower first quartile (43.5 weeks) and the highest variability (SD = 13.0), suggesting a greater concentration of early recidivism. Groups 4 and 5 display tighter distributions, though some variability remains. Group 6 (post-secondary education), despite its small size (n = 11), is notable for having all quartiles at the upper limit of the observation window (52 weeks) and the lowest standard deviation (SD = 3.02), which may reflect greater stability in outcomes. Group 2 (≤ Grade 6), also small (n = 24), presents low dispersion and a relatively high mean survival time, though interpretation should be made cautiously due to the limited number of observations.

Financial assistance appears to be associated with reduced vulnerability. Both groups share a median of 52 weeks, yet individuals who did not receive support display greater variability and more frequent early recidivism (Q1 = 44 weeks; minimum = 1 week). In contrast, those who received assistance have tighter distributions near the upper limit, suggesting a potential protective effect.

Marital status similarly plays a differentiating role. Married individuals show more clustered distributions at the 52-week mark and less variability (SD = 7.85), possibly reflecting the stabilizing influence of social support. Unmarried individuals have wider dispersion and earlier instances of reoffending.

Differences related to parole supervision are less pronounced. The median is identical across groups, but those not under parole have a slightly lower first quartile (48 weeks) and higher variability (SD = 13.2), hinting at marginally higher exposure to early recidivism.

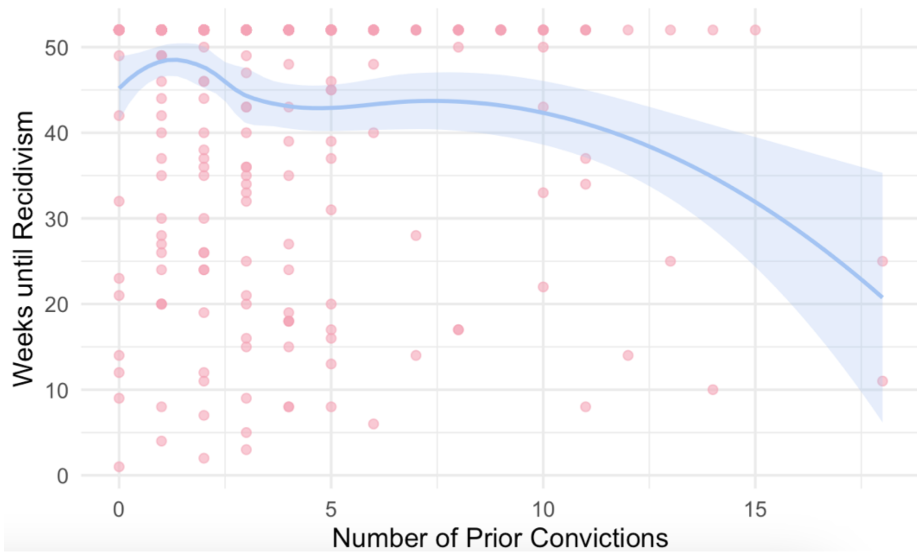
Compared to some of the more marked differences observed across other characteristics, variation by race appears more limited. Both subgroups share the same median and upper quartile (52 weeks), suggesting similar overall survival experiences. However, one group exhibits greater variability (SD = 13.0 compared to 9.73) and a slightly lower first quartile (49 versus 52 weeks). This may indicate greater heterogeneity in outcomes within that subgroup, without necessarily implying a direct association with recidivism.

Work experience stands out as one of the most powerful differentiators. Individuals with prior full-time employment show tight clustering around the 52-week mark and low variability (SD = 10.6). In contrast, those without such experience present a much broader distribution, with a markedly lower Q1 (36 weeks) and higher standard deviation (14.6), indicating increased susceptibility to early recidivism and emphasizing the protective role of labor market integration.

In addition to categorical variables, this section explores the relationship between the number of prior convictions and the timing of recidivism. Figure 2 illustrates this relationship using smoothed scatterplots with local polynomial regression (LOESS), which allows for the identification of non-linear trends without imposing a specific functional form.

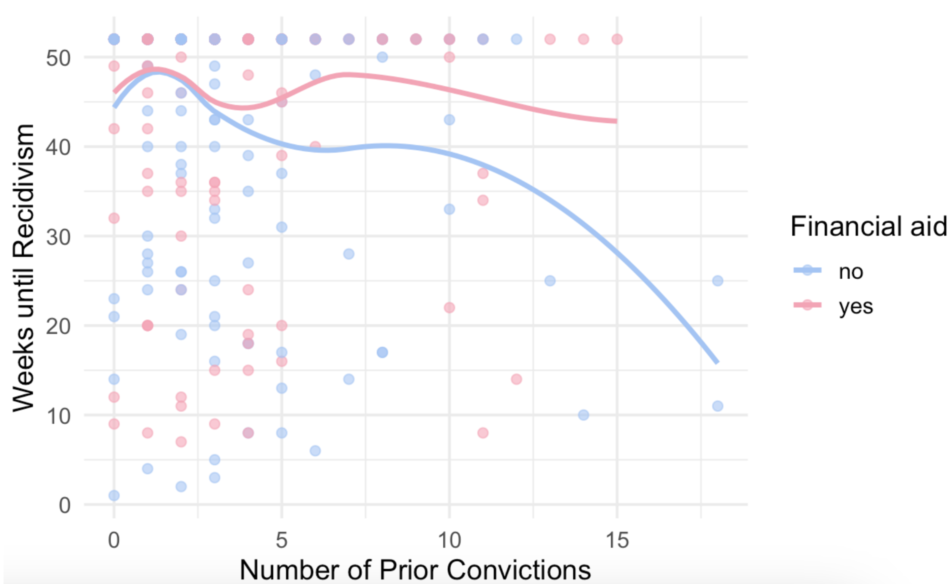
The results suggest a non-linear association. For individuals with relatively few prior convictions, the average time to recidivism remains high and stable, though individual variability persists. However, once the number of convictions reaches approximately 10 or more, a marked decline in non-recidivism duration is observed. This pattern points to a cumulative effect of criminal history, where repeated involvement with the justice system increases the likelihood of earlier reoffending.

Figure 2. Time to Recidivism by Number of Prior Convictions



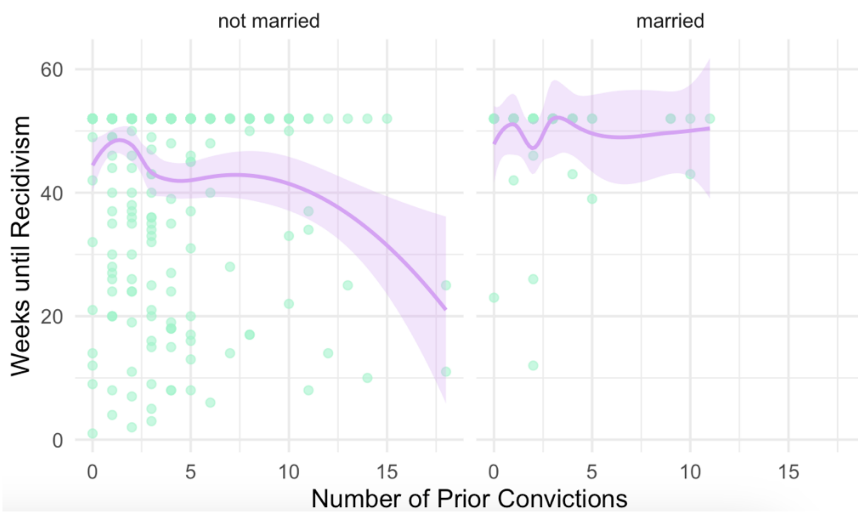
When financial assistance is considered (Figure 3), the differences become more pronounced. Individuals who did not receive economic support exhibit a consistent downward trend in the time to recidivism as their conviction history increases. In contrast, those who received financial assistance tend to remain offense-free for longer periods, even as their number of prior convictions rises. This reinforces the notion that targeted post-release support can mitigate the risks associated with more extensive criminal histories.

Figure 3. Recidivism by Convictions and Financial Support



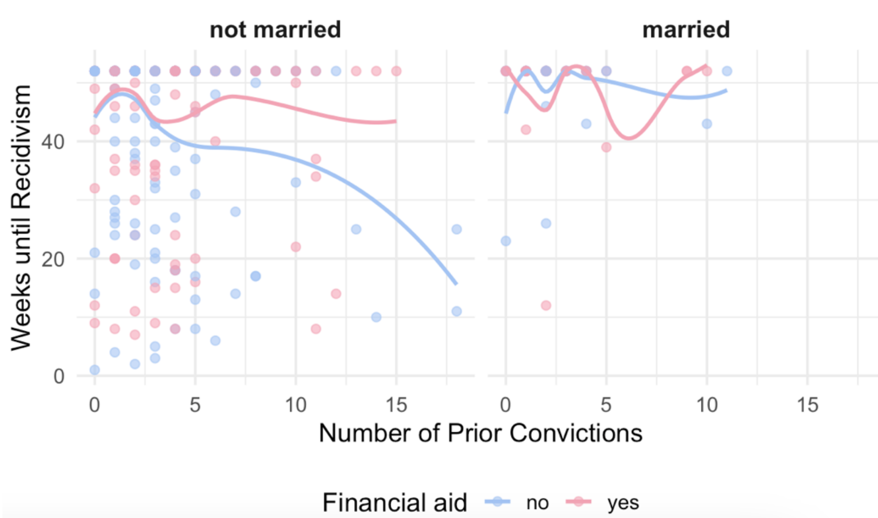
Marital status (Figure 4) also appears to moderate the relationship between prior convictions and recidivism. Among individuals who were not married at the time of release, a clear pattern emerges: the more prior convictions they have, the quicker they tend to reoffend. Conversely, married individuals maintain relatively stable and higher durations without reoffending, regardless of their criminal history. This finding suggests that marital ties may serve as a stabilizing factor and a form of social capital that supports reintegration

Figure 4. Recidivism by Convictions and Marital Status



Finally, when marital status and financial assistance (Figure 5) are considered together, the results provide further insight. Among unmarried individuals, those who did not receive financial assistance recidivate more quickly, particularly as prior convictions increase. In contrast, those with financial support experience longer non-recidivism periods. For married individuals, the differences between those with and without assistance are less pronounced, indicating that marital status alone may offer a substantial degree of protection. This highlights the potential for both financial and social support mechanisms to interact and strengthen post-release outcomes.

Figure 5. Recidivism by Financial Support and Marital Status



Log-Rank Test and Kaplan-Meier Curves

To complement the exploratory analysis, statistical tests were conducted to assess whether the time until recidivism differs significantly across subgroups. The log-rank test was applied to all categorical variables to determine whether the differences observed in the Kaplan-Meier curves are statistically meaningful.

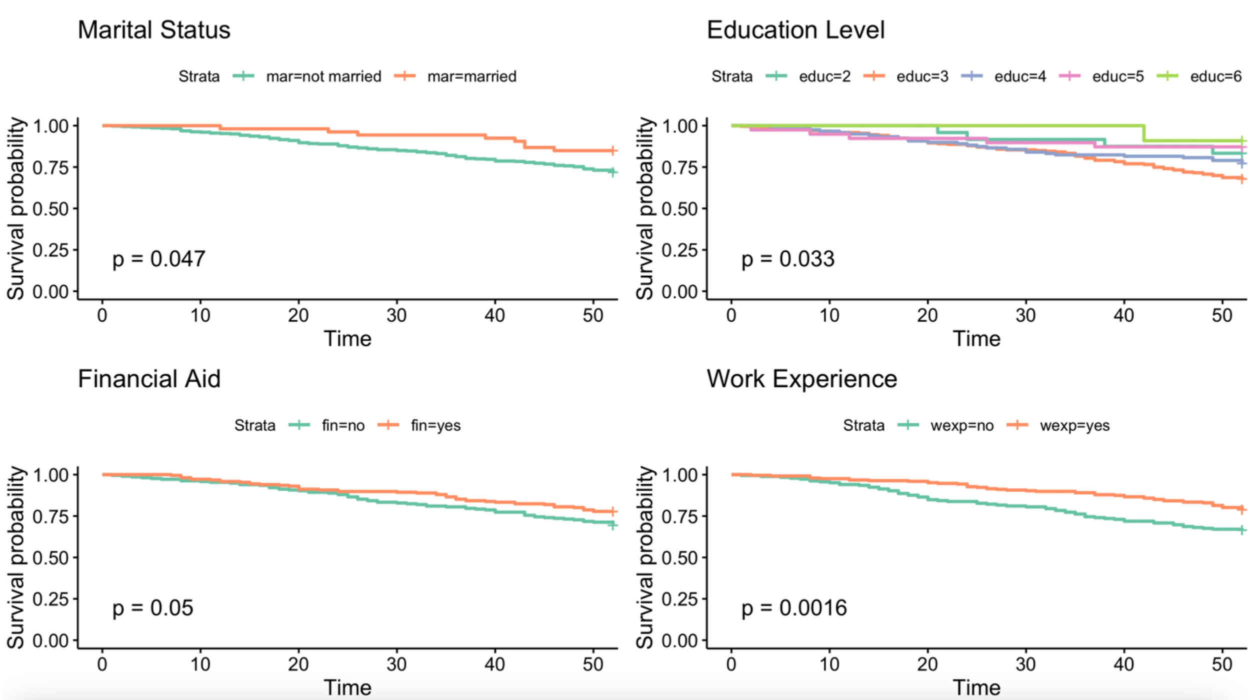
The results, summarized in Table 2, indicate that several individual characteristics are associated with distinct survival patterns. Work experience and education level emerged as the most discriminating variables, with p-values well below the conventional 0.05 threshold. Marital status and financial assistance also reached statistical significance, suggesting meaningful variation in the timing of recidivism. In contrast, race and parole status did not yield significant differences, indicating that these factors may not be strongly associated with the risk of reoffending during the one-year follow-up period.

Table 2: Log-Rank Test Results by Categorical Predictor

Variable	Race	Marital Status	Financial Support	Parole	Educational Level	Work Experience
p-value	0.44	0.05	0.05	0.57	0.03	0.00

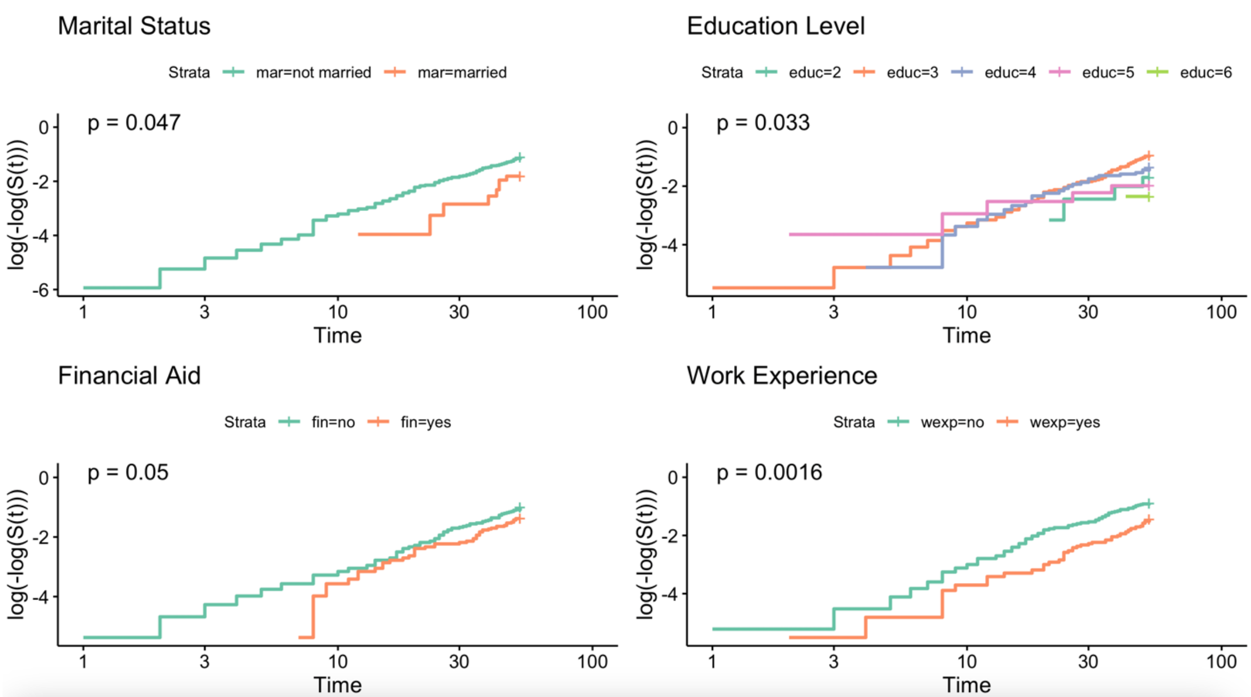
To visualize the group differences identified by the log-rank test, Kaplan-Meier survival curves were estimated for the covariates with p-values less than or equal to 0.05. These non-parametric curves provide a visual summary of the probability of remaining offense-free over time across subgroups. As shown in Figure 6, survival probabilities are consistently higher for individuals with protective characteristics such as being married, having prior work experience, or receiving financial assistance.

Figure 6. Kaplan-Meier Survival Curves by Individual Characteristics



While Kaplan-Meier curves are useful for identifying survival differences, they do not assess whether the proportional hazards assumption—a requirement for the Cox regression model—is met. To address this, complementary log-log plots were generated for the same variables. These plots, presented in Figure 7, offer a visual diagnostic of the proportionality of hazards by displaying log(-log(S(t))) transformations over time.

Figure 7. Complementary Log-Log Curves by Individual Characteristics



The plots suggest that, for marital status, financial assistance, and work experience, the group-specific curves remain relatively parallel over time and do not intersect sharply—patterns that are consistent with the proportional hazards assumption. However, the plot for education level shows greater variability and some curve crossings, especially among intermediate categories, which may indicate a potential violation of the assumption for this variable. These visual insights highlight the importance of carefully interpreting the role of education in the multivariate context and considering robustness checks or alternative specifications where needed.

Cox Proportional Hazards Model

To better understand the factors that influence the timing of recidivism, a Cox proportional hazards model was estimated using all available predictors. This approach allows for the simultaneous evaluation of multiple variables, offering insights into whether specific characteristics are associated with an increased or decreased likelihood of reoffending over time.

The full model included financial assistance, race, work experience, marital status, parole status, number of prior convictions, and education level. As shown in Table 3, two variables were found to be statistically significant. Receiving financial assistance after release was associated with a substantially lower risk of recidivism. Specifically, individuals who received economic support were 36 percent less likely to reoffend compared to those who did not receive such assistance (hazard ratio = 0.64; p = 0.021). Additionally, the number of prior convictions emerged as a strong predictor of earlier reoffending. Each additional conviction increased the hazard of recidivism by approximately 8 percent (hazard ratio = 1.08; p = 0.009).

Other factors—including work experience, marital status, education level, race, and parole status—showed directional trends consistent with the existing literature, suggesting potential protective or risk effects. However, these coefficients did not reach conventional levels of statistical significance in this model. For example, both marital status and work experience appeared to be associated with a reduced hazard of reoffending, but the evidence was not strong enough to confirm these associations in this particular dataset.

Table 3. Cox Model – Full Specification

Predictor	HR	Coefficient	Std. Error	z-value	p-value
Financial Support ("Yes")	-0.4429	0.6422	0.1921	-2.3060	0.0211*
Race ("Other")	-0.3156	0.7293	0.3115	-1.0130	0.3108
Work Experience ("Yes")	-0.2911	0.7474	0.2048	-1.4220	0.1552
Marital Status ("Married")	-0.5379	0.5840	0.3795	-1.4180	0.1563
Parole ("Yes")	-0.0439	0.9571	0.1942	-0.2260	0.8213
Prior Convictions (numeric)	0.0756	1.0785	0.0290	2.6080	0.0091**
Education Level (3)	0.7616	2.1416	0.5157	1.4770	0.1397
Education Level (4)	0.5129	1.6701	0.5387	0.9520	0.3411
Education Level (5)	-0.0250	0.9753	0.6732	-0.0370	0.9704
Education Level (6)	-0.3921	0.6756	1.1220	-0.3490	0.7268
Number of observations = 432					
Number of events = 114					

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Given these results, a reduced version of the model was estimated, including only the two predictors that showed statistically significant effects in the full specification: financial assistance and number of prior convictions. The findings from this simplified model, presented in Table 4, were consistent with those from the full model. Financial assistance retained its protective role, with recipients being 33 percent less likely to reoffend than non-recipients (hazard ratio = 0.67; p = 0.036). The effect of prior convictions also remained robust and slightly stronger in this specification. Each additional prior conviction was associated with an 11 percent increase in the likelihood of recidivism (hazard ratio = 1.11; $p < 0.001$).

Table 4. Cox Model – Reduced Specification

	HR	Coefficient	Std. Error	z-value	p-value
Financial Support ("Yes")	-0.3988	0.6711	0.1901	-2.0980	0.0359*
Prior Convictions	0.1041	1.1097	0.0268	3.8920	0.0001***
Number of observations = 432					
Number of events = 114					

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Testing the Proportional Hazards Assumption

Before drawing conclusions from these results, it is essential to verify that the key assumption underpinning the Cox model—namely, the proportional hazards assumption—is satisfied. This assumption holds that the effect of each predictor remains constant over time. In other words, if a variable reduces the risk of recidivism at the beginning of the follow-up period, it is assumed to reduce that risk by the same proportion throughout the entire duration. Violations of this assumption could undermine the reliability of the model's estimates and, by extension, their policy relevance.

To formally assess this assumption, the Schoenfeld residuals test was applied. This test evaluates whether the residuals from the Cox model are correlated with time, which would indicate that the effect of a predictor on the hazard of recidivism may change over the course of the observation period. The results are presented in Table 5.

The p-values for both financial assistance and number of prior convictions are well above the 0.05 threshold, indicating no evidence that their effects change over time. Moreover, the global test confirms that the model satisfies the proportional hazards assumption. This reinforces confidence in the Cox regression results, particularly the lasting protective role of financial assistance and the heightened risk posed by repeated convictions.

Table 5. Schoenfeld Test – Proportional Hazards Assumption

	Chi-square	Degree of Freedom	p-value
Financial Support ("Yes")	0.0741	1	0.79
Prior Convictions	0.5093	1	0.48
Global Test	0.5654	2	0.75

Conclusion

This analysis examined the timing of recidivism among individuals recently released from prison, using survival analysis techniques to identify the characteristics most strongly associated with the risk of reoffending. The results consistently highlight two key predictors: financial assistance and prior criminal history. Individuals who received economic support were significantly less likely to recidivate, suggesting that financial support plays a protective role after release. Conversely, a higher number of prior convictions was strongly associated with earlier and more frequent recidivism, reinforcing the cumulative impact of criminal history on future outcomes.

Diagnostic tests confirmed that the proportional hazards assumption holds for the final Cox model, lending credibility to the robustness of these findings. The stability of these effects over time strengthens the interpretation of the estimated relationships and supports the overall validity of the survival analysis results.

Together, these findings contribute to a better understanding of how individual characteristics relate to the timing of recidivism and demonstrate the usefulness of survival analysis for studying dynamic post-release outcomes.