

Waiting for Trial: A Case Study of Detention Times Prior to Sentencing

Lillian Clark

12/7/2021

1. Introduction

1.1 Context

In 2019, 23.3% of incarcerated people in the United States were legally innocent – they had not been tried on their charges in a court of law.¹ The Prison Policy Initiative found that the number of people in jail pretrial has nearly quadrupled since the 1980s.² Pretrial detention is severely destabilizing for incarcerated individuals and their families. Even a few days in jail can cause people to lose their jobs, their homes, or custody of their children. Studies have shown that people who are detained pretrial are more likely to be convicted and more likely to receive harsher sentences than those who go free.³ Pretrial detention is also associated with significantly higher recidivism rates.⁴

Widespread use of money bail has led to the criminalization of poverty, where individuals who can afford to pay a bail bondsman's fee or to tie up thousands of their own dollars in the court system for months or years can go free, and individuals without those funds sit in jail. This can force innocent individuals who can't pay to claim guilt and accept a plea deal simply to get out of jail faster. In the U.S., there are also clear racial disparities among pretrial detainees. Young Black men are roughly 50% more likely to be detained than white individuals.⁵ Some people are held in limbo by the criminal legal system for amounts of time comparable to what they would serve if they were convicted. In 2017, The New York Times profiled Kharon Davis, a man who spent at least ten years in county jail awaiting trial. They wrote, "Though he has not been found guilty, Mr. Davis has already served half of the minimum sentence for murder."⁶

Not everyone held in jail is awaiting trial – there are also individuals with relatively minor convictions like parole violations or misdemeanors who serve their full sentences there, usually for less than a year. The proportion of pretrial detainees in jail populations has increased from 53% in 1970 to 64% in 2015.⁷ Some data is available for web scraping from local jails but there are significant problems with transparency, inaccuracies or data entry issues in this data. (INSERT SCREENSHOT OF DURHAM JAIL LIST) These problems make it difficult to determine which jail inhabitants have been convicted and which have not. It also makes studying pretrial detention and inequities harder.⁸

The ICPSR 2016 Survey of Prison Inmates contains data from 364 prisons across the United States. 24,848 incarcerated individuals participated, 81% of which were held in state prisons and the other 19% in federal prisons. Part of the survey asked respondents how long they had been detained in jail.

¹World Prison Brief <https://www.prisonstudies.org/country/united-states-america>

²https://www.prisonpolicy.org/research/pretrial_detention/

³need citation

⁴Citation

⁵Prison Policy Initiative

⁶<https://www.nytimes.com/2017/09/19/us/alabama-kharon-davis-speedy.html>

⁷Vera Brief <https://www.vera.org/downloads/publications/Justice-Denied-Evidence-Brief.pdf>

⁸The Transparency of Jail Data, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3741638

- Error involved in any kind of survey data: jail time is self-reported and not verified from a second source. Also, the survey question doesn't ask participants to specify whether this jail time was pre-detention. Using jail time as length of pretrial detention is not a perfect proxy, but we have reason to believe it is a good one (at least given the way the system is supposed to work), since those serving their sentences for minor convictions in jails should not have been surveyed and split jail-prison sentences are not the norm – generally, individuals in prison, if they were detained pretrial, did so in jails, and were transferred to prisons after their convictions.
- 391 individuals, or 1.6% of all observations in the data, reported that they had not yet been sentenced. These pretrial wait times are right-censored, but the group is very small compared to the overall survey participants.
- This data is restricted to individuals in prison, so will exclude those with dismissed or not guilty charges who were also held pretrial. However, it still provides an important window into the punishment people face in America before they are even tried (the wait for justice).

Research questions:

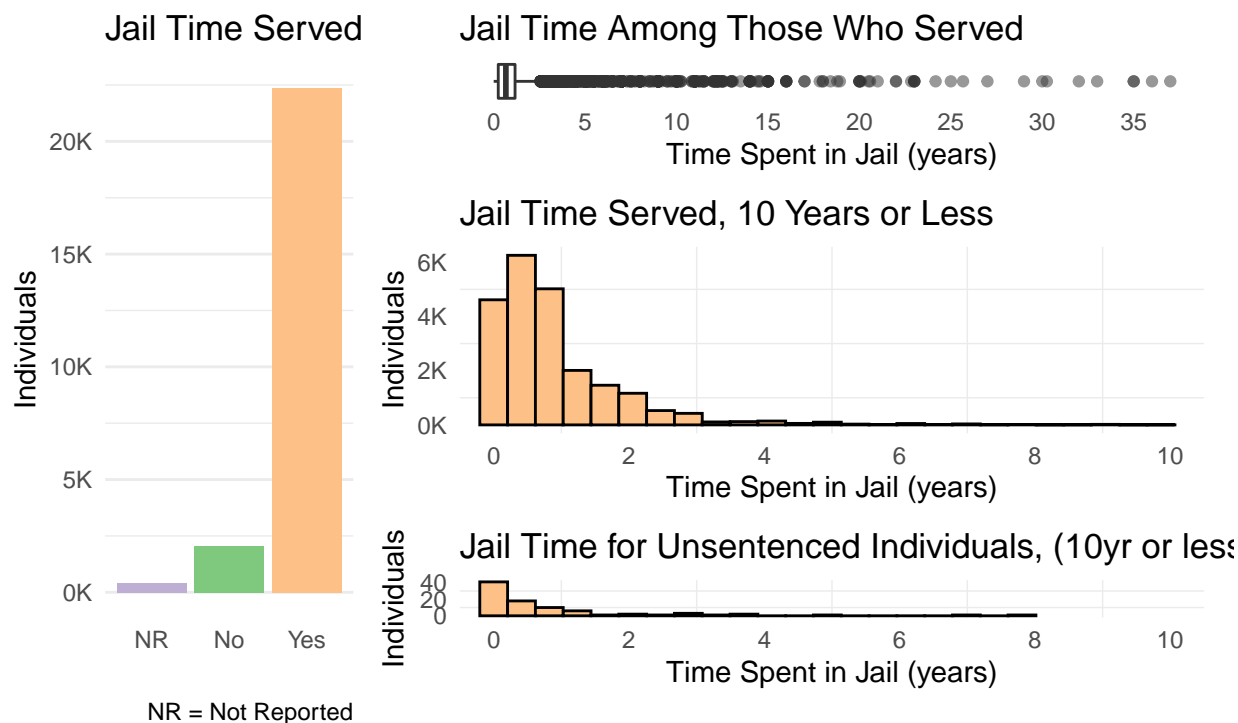
- 1) When controlling for aspects of the main offense they are prosecuted for and other policy-related covariates, do individuals incarcerated in United States prisons in different racial and ethnic groups wait longer before sentencing?
- 2) Does time detained pretrial have any relationship with characteristics, such as sex, race, age, and whether they were known by the detained individual, of the victims of “violent” crimes?

Extended pretrial limbo impacts both the detained accused and victims waiting for resolution. Looking at victim characteristics is essential because of the way racial bias and violence have historically been and continue to be codified in the criminal legal system, where white victims are more likely to get attention and justice than Black and brown victims. We see this in the disproportionate national media fervor over cases of white women who go missing or face violence while missing Black and Indigenous women are often ignored or overlooked. Class can play into this as well, with lower-class white victims experiencing different levels of treatment and attention than their upper-class counterparts. Unfortunately the ICPSR survey does not include information on victim income, so I lack the data necessary to add this consideration to analysis. Regardless, the disparities in treatment based on racial dynamics are complicated. It's difficult to outline an ideal outcome because we can imagine scenarios in which less time spent pretrial with this dynamic might be worse for the defendant. Like in the case of Walter McMillian⁹, innocent Black men have frequently been accused of violence against white women and swept through a speedy and unfair trial on a wave of white supremacist fervor. Hasty prosecution and extensive pretrial detention are two very important ways justice is denied, but they manifest as opposite effects in our data. This analysis is limited in its ability to address all that complexity, keeping in mind the meanings of such varied outcomes for both victims and defendants alike. However, it will attempt to shed some light on whether there are differences based on victim characteristics and what the magnitudes of those differences are.

1.2 Data Description

Figure 2: Distribution of Jail Time

⁹Bryan Stevenson, *Just Mercy*



In the data, 90.0% of individuals reported serving jail time, 8.3% reported serving no jail time, and 1.7% skipped the related survey questions (jail time served and amount of jail time) (CHECK THIS). FIX FIGURE 457 individuals have missing data for jail time, and INSERT JUSTIFICATION these individuals will be excluded from the primary and secondary analyses (after the jail time served logistic model).

The maximum jail time reported in the data was 37 years, though the majority of individuals served 3 years or less. In 2017, The New York Times wrote that Kharon Davis's 10-year wait pretrial was "among the most protracted" they could find.¹⁰ Thus, this will serve as a best estimate for expected maximum jail time, and we will filter the data before analysis to remove observations with jail time greater than or equal to 10 years. This removes 285 observations or 1.2% of the dataset FIX FIGURE. Also, three individuals who did not report their military status and ten individuals who did not report their sentencing status were excluded from analysis.

ELABORATE ON MISSINGNESS DISCUSSION (from C3 From visualizations of missingness in Appendix section A, we found income seems to have the highest percentage of missingness (14%), followed by years in Durham (3.5%). Based on background knowledge, we assumed the data is **missing at random (MAR)**. We believe the observed missingness can be largely explained by other explanatory variables which we have data for rather than unobserved or unmeasured variables.)

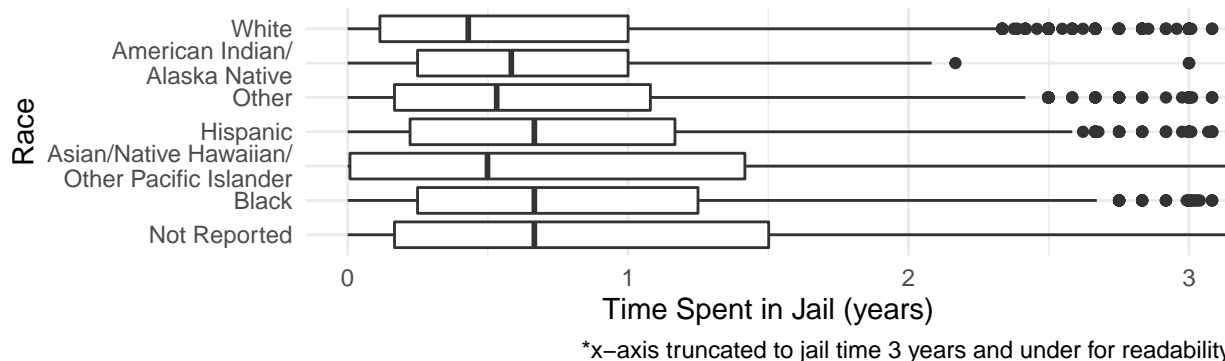
Additionally, incarcerated individuals were asked whether they had been sentenced, and a small number FIX FIGURE (133 individuals, or 0.55% of the data after the above filtering) reported that they had not yet been sentenced. Thus, their pretrial detention times are right-censored. The vast majority of these unsentenced individuals were arrested in more recent years, between 2013 and 2016.

We are interested in controlling for a number of covariates when examining discrepancies in pretrial detention. These can be split into three groups: **policy-related** covariates (arrest year, authority detained by, and state of detention facility), **charge-related** covariates (primary offense and primary offense type, legal status at arrest, possession of firearm at offense, under the influence of alcohol at offense, and under the influence of drugs at offense), and **demographic-related** covariates (age at arrest, race, sex, citizenship, veteran status, education level, and whether homeless in 12 months prior to offense). Policy-related covariates matter because of changes in pretrial detention policy over time (such as during the "tough on crime" era in the 1980s and early 1990s) and across states and jurisdictions. The charge-related covariates represent all the

¹⁰citation

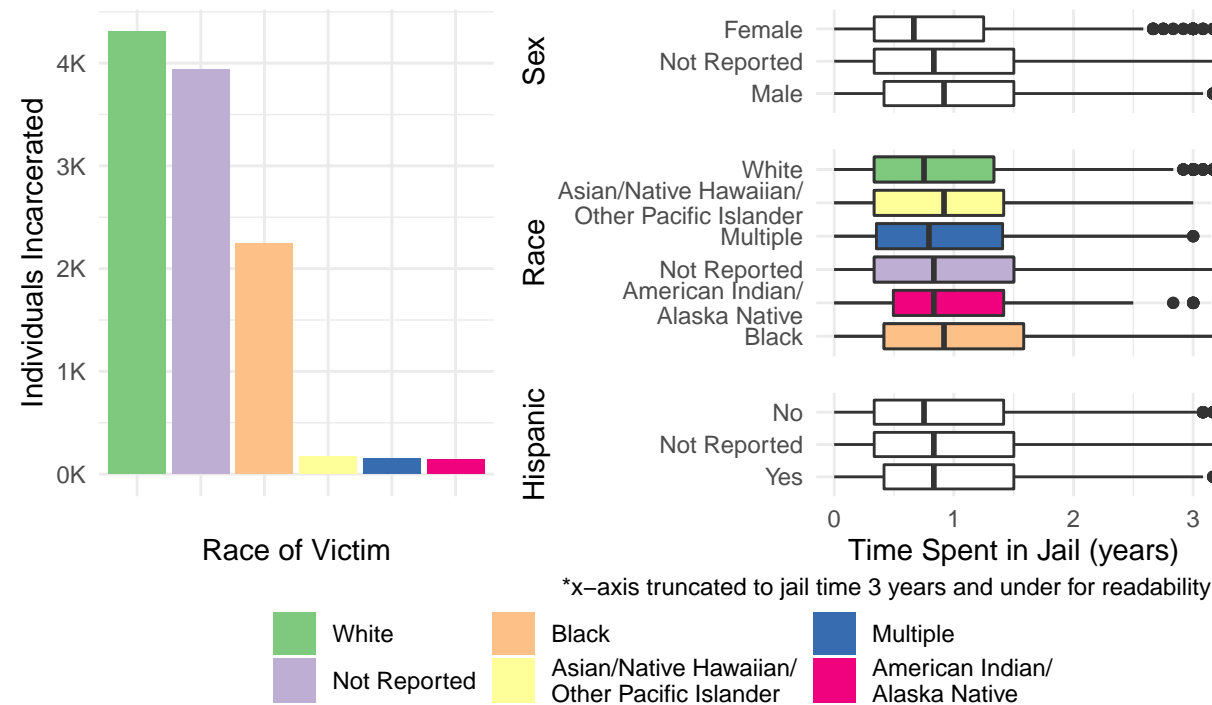
offense-related information available in the ICPSR data, information which would have been available to the district attorneys and judges overseeing these cases and making decisions about bond and detention. And demographic-related covariates allow us to look for discrepancies in treatment between groups of people.

Figure 3: Distribution of Jail Time by Race of Incarcerated Individual



Before controlling for covariates, mean jail time seems lowest for white incarcerated individuals and highest for Black individuals and those who did not report their race. The discrepancies between groups in Figure 3 may not appear large. However, in a world where just days or weeks in jail can make a difference in the resources and support networks available to incarcerated individuals, small discrepancies matter.

Figure 4: Distribution of Jail Time for “Violent” Offenses by Victim Characteristics



Data for our secondary analysis contains only individuals serving time for a primary offense labeled ‘violent.’ Victim information for those convicted of drug and property crimes, offenses which are often victimless or which lack easily identifiable victims, is largely missing. However, 64% of those with a violent primary offense did report victim race. Missingness patterns in victim sex and Hispanic ethnicity were similar. In Figure 4, we observe that individuals convicted of crimes against male victims appear to have a longer mean detention period pretrial. Black and Indigenous victims see the highest mean pretrial detention; white and multiracial victims the lowest. Hispanic victims also appear to see higher mean jail time than non-Hispanic victims.

1.3 Hypothesis

Due to the research cited above and preliminary data visualization, we hypothesize that Black, Indigenous and Latinx incarcerated people serve longer jail times. Also, despite the punishing and justice-denying effects of extended pretrial incarceration for those facing charges, due to the imbalance of community attention and outrage surrounding (especially female) white victims versus victims of color, we hypothesize that violent crimes committed against white victims will be associated with shorter detention times pretrial compared to Black and Indigenous victims.

2. Methodology

2.1 Preliminary Analysis: Any Jail Time Served

Our first goal is to understand whether jail time served is associated with the race of the incarcerated individual. While survival analysis will allow for incorporation of individuals who have not yet been sentenced into our model, roughly 8% of the data consists of people who reported serving no jail time and would have survival time 0. These individuals are important to consider. They likely represent a group that could afford to post bail or pay a bail bondsman, or that live in states with pretrial detention policies that don't heavily rely on money bail, places like New Jersey and Washington D.C..¹¹ To consider those who were not detained pretrial before removing them from the data for survival analysis, we will first conduct multinomial logistic regression with the jail time served indicator as our response and policy, charge, and demographic covariates as predictors.

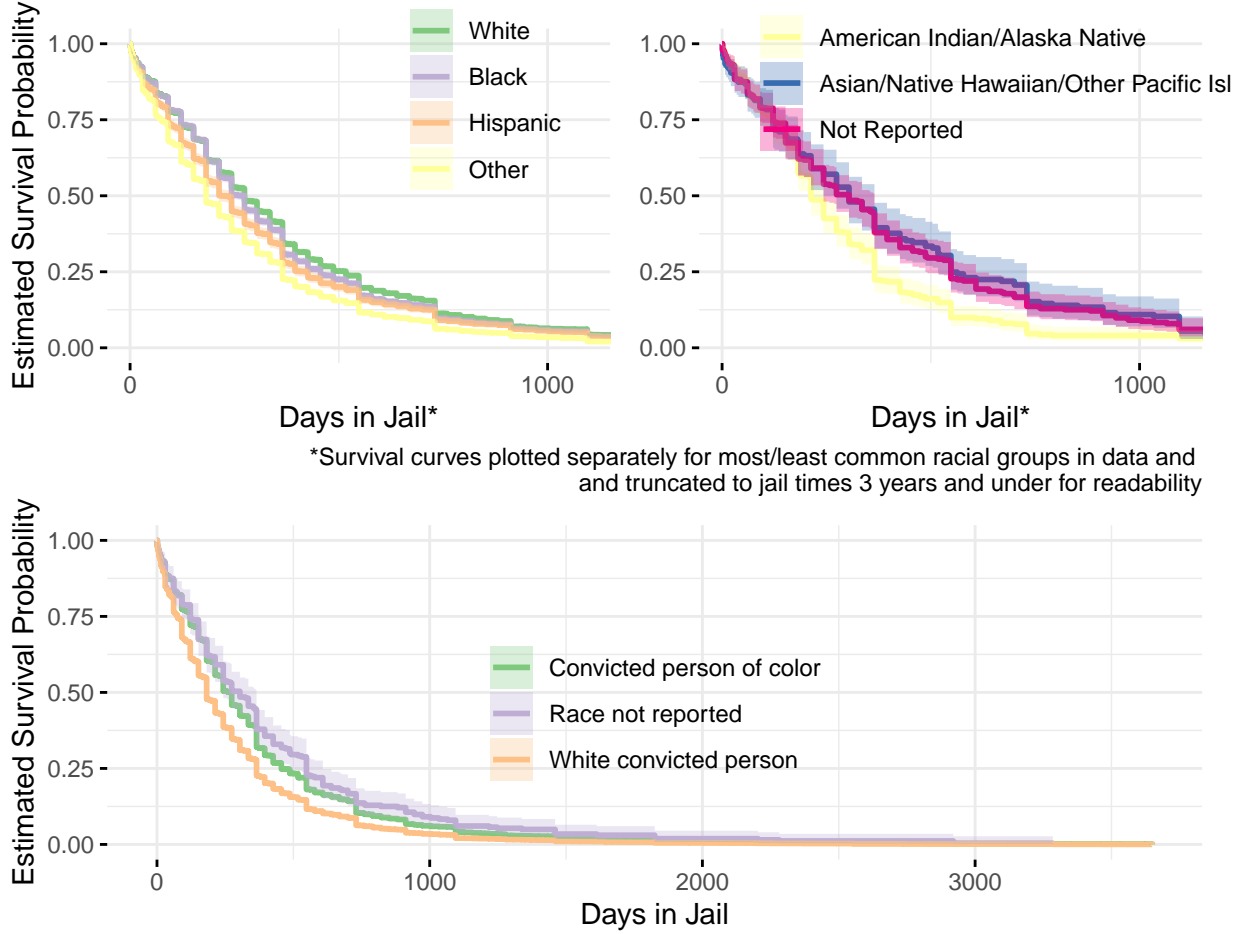
We opt for multinomial logistic regression over binary logistic regression with only jail time served/not served as outcome in order to account for potential meaning in missingness in the data. Instead of filtering out those with jail time not reported, we consider "Not Reported" as a third outcome in our model. TALK ABOUT ASSUMPTIONS

2.2 Primary Analysis: Racial Discrepancies in Jail Time

Our next goal is to observe whether an individual's race is associated with jail time, and if an association does exist, to understand the effect size of race on days in pretrial detention.

Figure n: Survival Curves for Race of Incarcerated Individual

¹¹citation



We plot survival curves for the data from a non-parametric Kaplan-Meier estimate. When racial categories are simplified to white and non-white, as shown at the bottom of Figure n ~~FIX THIS~~, the curves for different groups do not cross. They also remain roughly parallel at most survival times. This provides empirical evidence of proportional hazard and establishes the power of a log-rank test as a method to compare groups. Such a test of survival times for white incarcerated individuals, non-white individuals, and those who did not report their race has significant p-value < 0.001 (see Appendix B ~~FIX THIS~~ for full test). We will further explore this racial discrepancy with a survival model.

To further examine the scale of this difference and incorporate confounders, we built an accelerated failure time (AFT) model with error term ϵ_i normally distributed and survival time T_i log-normally distributed, specified as follows:

$$\log(T_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \epsilon_i$$

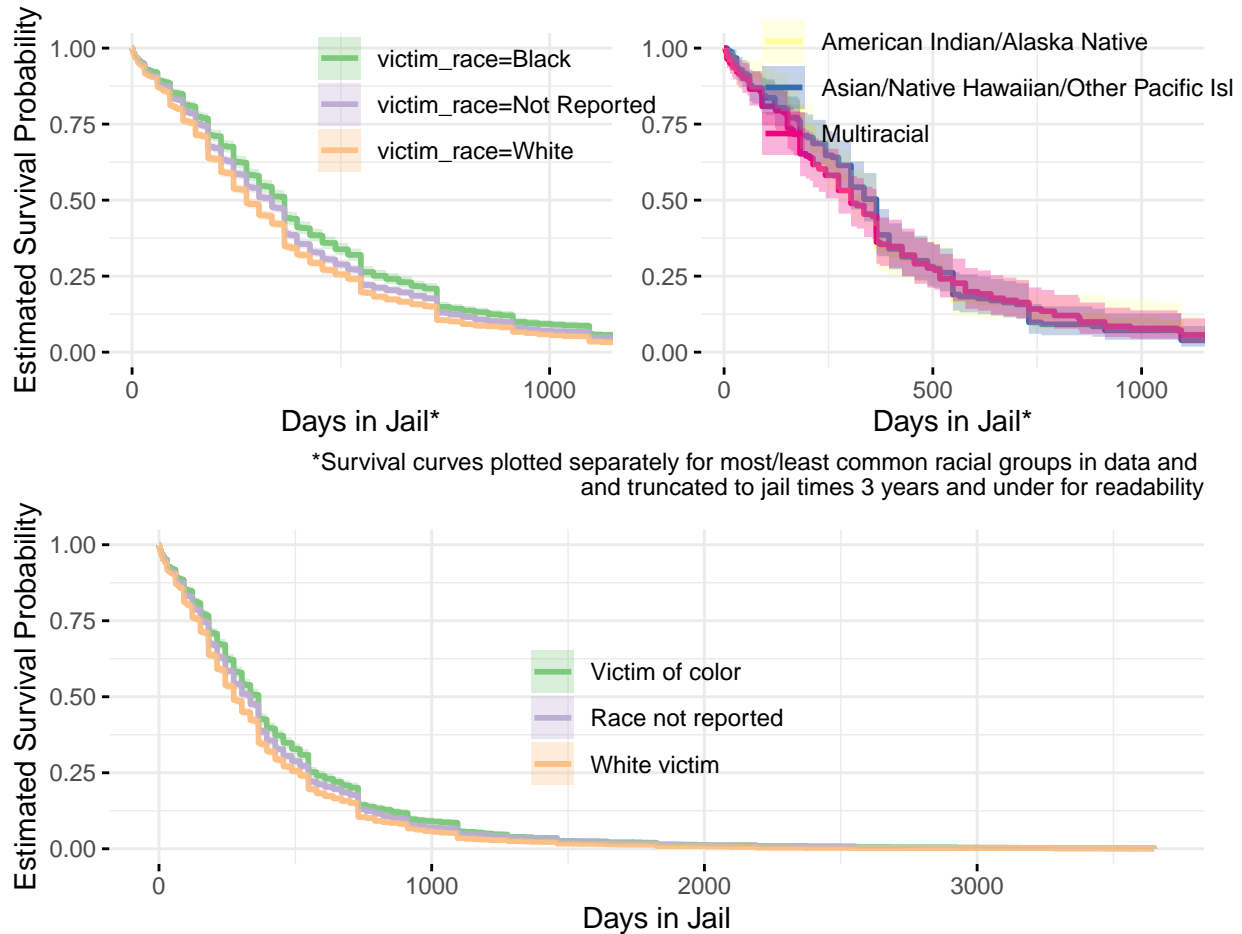
We included the following covariates: primary offense, legal status at arrest, firearm present at offense, under the influence of alcohol at offense, under the influence of drugs at offense, arrest year range, authority held by, state of incarceration, age at arrest, race, sex, citizenship status, veteran status, education level, and whether homeless in the year prior to offense. We compared AFT models with Weibull, log-normal, log-logistic, and exponential hazard distributions in Appendix ~~FIX THIS~~ and conducted model selection with Kaplan-Meier residual plots under the survival functions of these distributions. These plots show the log-normal model performs well, that the curve of its distribution fits the residuals most closely and satisfies assumptions (see Appendix ~~FIX THIS~~). As shown in the survival curves above, there are several different ways race is encoded in the data, either in 1) the groups white, Black, hispanic, other, American Indian/Alaska Native, Asian/Native Hawaiian/Other Pacific Islander, and not reported; 2) the groups white, non-white, and not

reported; and 3) the groups Black, non-Black, and not reported. After model selection, we conducted variable selection via Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine which among these three variables to include as a confounder in our log-normal model. Both metrics select the model which labels individuals as white/non-white.

In lieu of an AFT model, we could have pursued a Cox proportional hazards model. Although a semi-parametric model may have done better to capture the complexity in our data, a proportional hazards model would not do well to estimate the magnitude of the difference in jail time served between groups. Our goal is to ascertain how long individuals spend detained pretrial, not their instantaneous hazard rate of sentencing. Therefore, an AFT model is more appropriate for our analysis. Estimating survival times is possible with a Cox proportional hazards model through Breslow estimator calculations, but AFT model coefficients provide a more straightforward approach. CHECK THIS Our two assumptions in this approach are that individuals with censored jail times have the same survival prospects as those without censoring at every point in time and that survival probability functions have remained roughly constant over the period of data collection.

2.3 Secondary Analysis: Jail Time by Victim Characteristics

Figure n: Title



Our methodology for our secondary analysis of victim characteristics is identical to the one above, with the exception of the initial inclusion of victim race, victim sex, victim age, victim hispanic ethnicity and whether the victim was known by the individual as additional covariates. Our interpretation goals remain oriented on survival times, not hazards, so we build AFT models to compare. We find a log-normal AFT model to be the best fit when its hazard distribution is compared to a plot of the Kaplan-Meier residuals, and comparing AIC and BIC leads us to select victim race white/non-white as our method of encoding race.

3. Results

3.1 Preliminary Analysis

Table 1: Multinomial Logistic Model Output

Y Level	Model Term	Estimate	Standard Error	Test Statistic	p-Value
Not Reported	raceBlack	-1.289	0.586	-2.201	0.028
Not Reported	raceOther	-1.803	0.642	-2.808	0.005
Not Reported	raceWhite	-1.457	0.582	-2.504	0.012
Not Reported	sexMale	0.587	0.279	2.101	0.036
Not Reported	educationLess Than High School	1.370	0.562	2.435	0.015
Not Reported	homeless_12mo_priorYes	0.723	0.311	2.322	0.020
Yes	age_at_arrest36-50	-0.125	0.060	-2.091	0.037
Yes	age_at_arrestOver 50	-0.526	0.087	-6.016	<0.001
Yes	sexMale	0.235	0.063	3.721	<0.001
Yes	citizenNon-citizen	0.471	0.120	3.909	<0.001
Yes	educationHigh School Graduate	0.485	0.096	5.024	<0.001
Yes	educationLess Than High School	0.744	0.094	7.941	<0.001
Yes	educationSome College	0.466	0.100	4.645	<0.001
Yes	homeless_12mo_priorYes	0.439	0.108	4.074	<0.001

3.2 Primary Analysis

Table 2: AFT (log-normal) Model Output (with race white/non-white)

Model Term	Estimate	Standard Error	Test Statistic	p-Value
age_at_arrest36-50	-0.049	0.021	-2.338	0.019
age_at_arrestOver 50	-0.156	0.038	-4.159	<0.001
race_whiteY	-0.047	0.020	-2.296	0.022
sexMale	0.226	0.023	9.834	<0.001
citizenNon-citizen	0.314	0.041	7.682	<0.001
educationHigh School Graduate	0.173	0.046	3.756	<0.001
educationLess Than High School	0.249	0.045	5.600	<0.001

3.3 Secondary Analysis

Table 3: AFT (log-normal) Model Output (with race white/non-white)

Model Term	Estimate	Standard Error	Test Statistic	p-Value
age_at_arrestOver 50	-0.134	0.054	-2.463	0.014
age_at_arrestUnder 18	0.111	0.052	2.138	0.033
race_whiteY	-0.117	0.030	-3.973	<0.001
sexMale	0.169	0.034	4.945	<0.001
educationHigh School Graduate	0.157	0.064	2.451	0.014
educationLess Than High School	0.227	0.063	3.625	<0.001
homeless_12mo_priorYes	0.081	0.041	1.982	0.048
victim_ageUnder 12 Years	0.119	0.058	2.060	0.039

4. Discussion

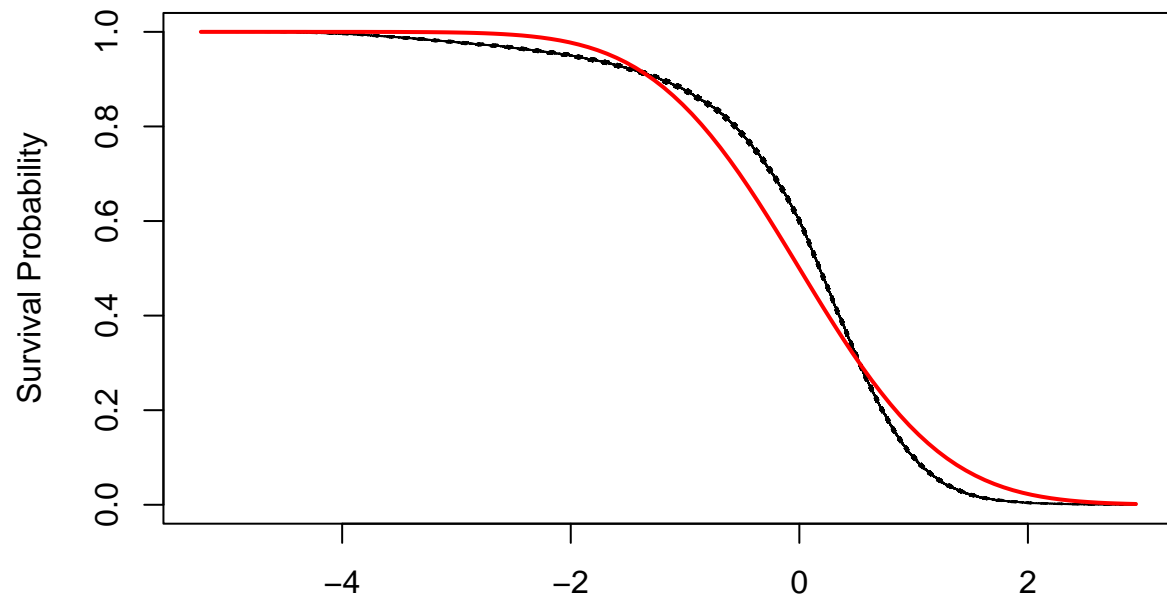
4.1 Conclusions

4.2 Limitations and Future Directions

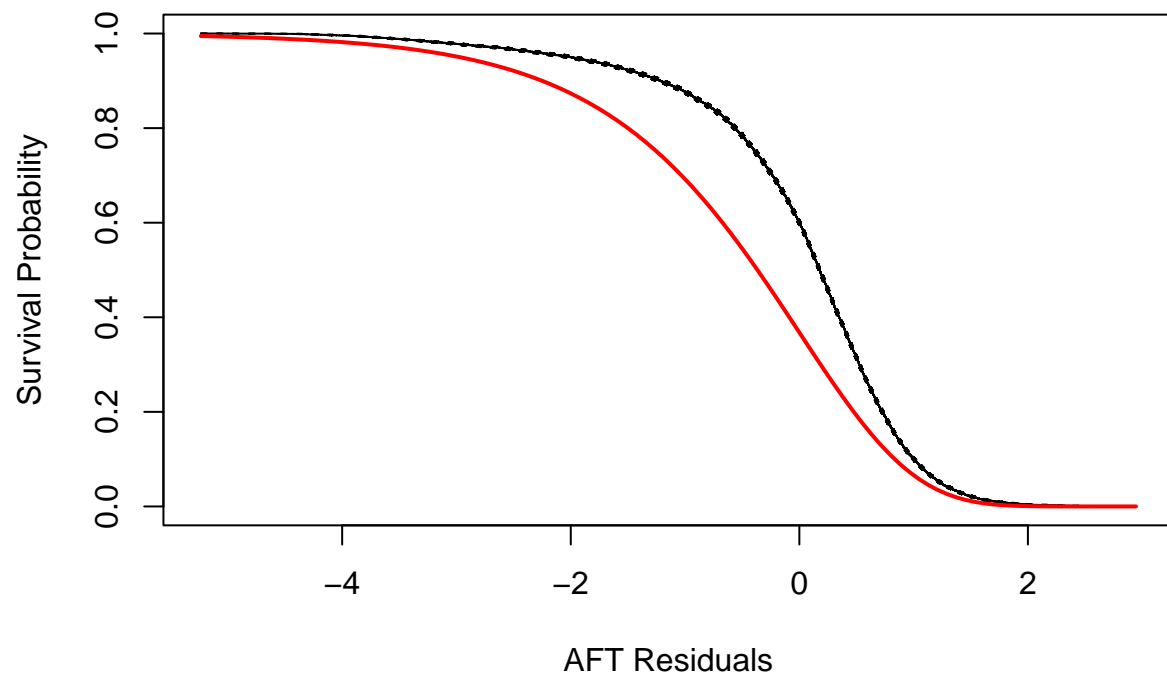
4.3 Summary

Appendix

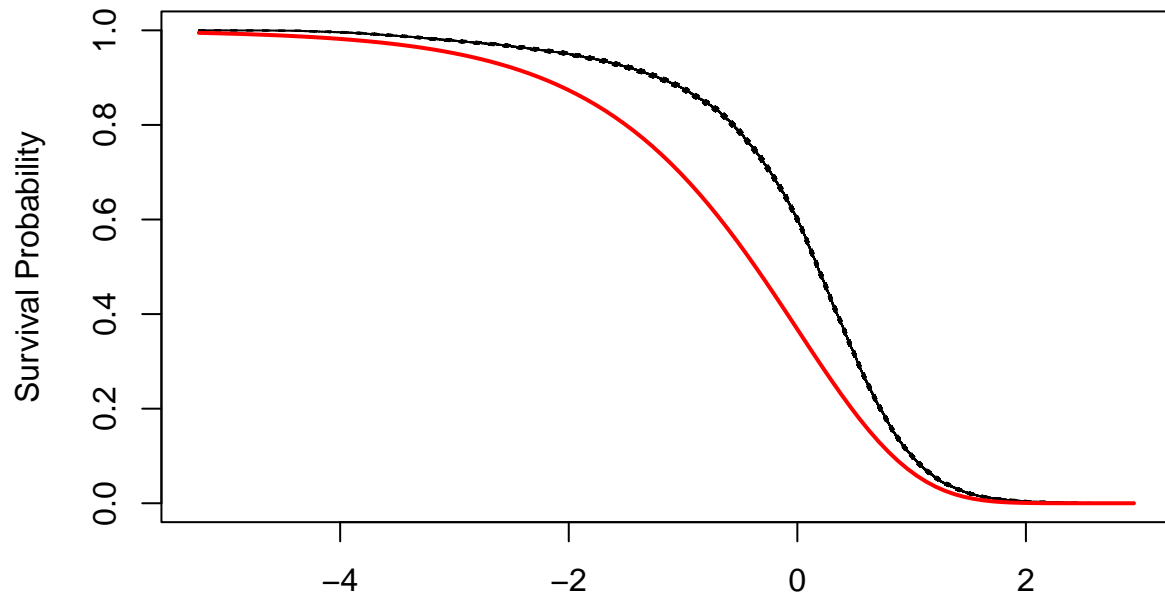
Log-Normal Model with All Confounders



Exponential Model with All Confounders



Weibull Model with All Confounders



Logistic Model with All Confounders

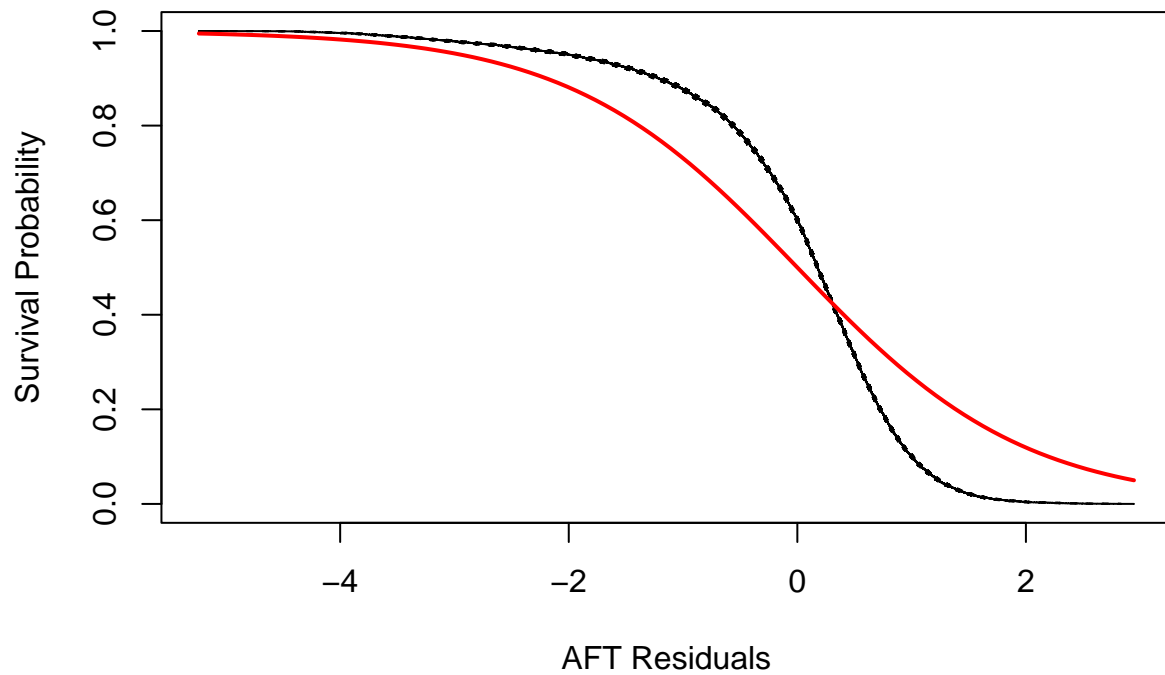
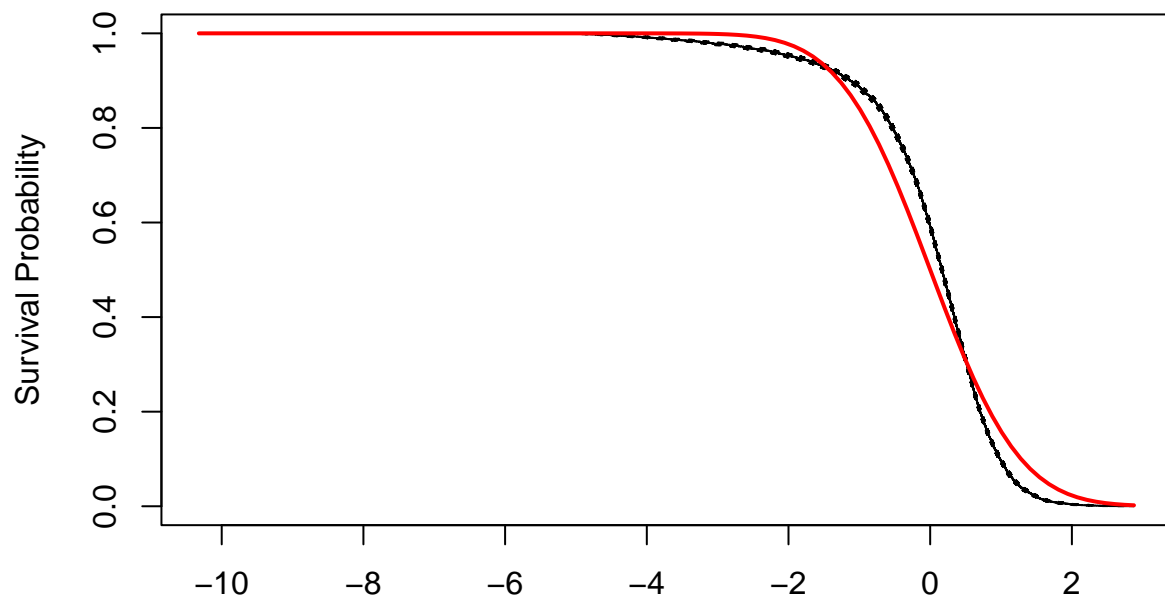


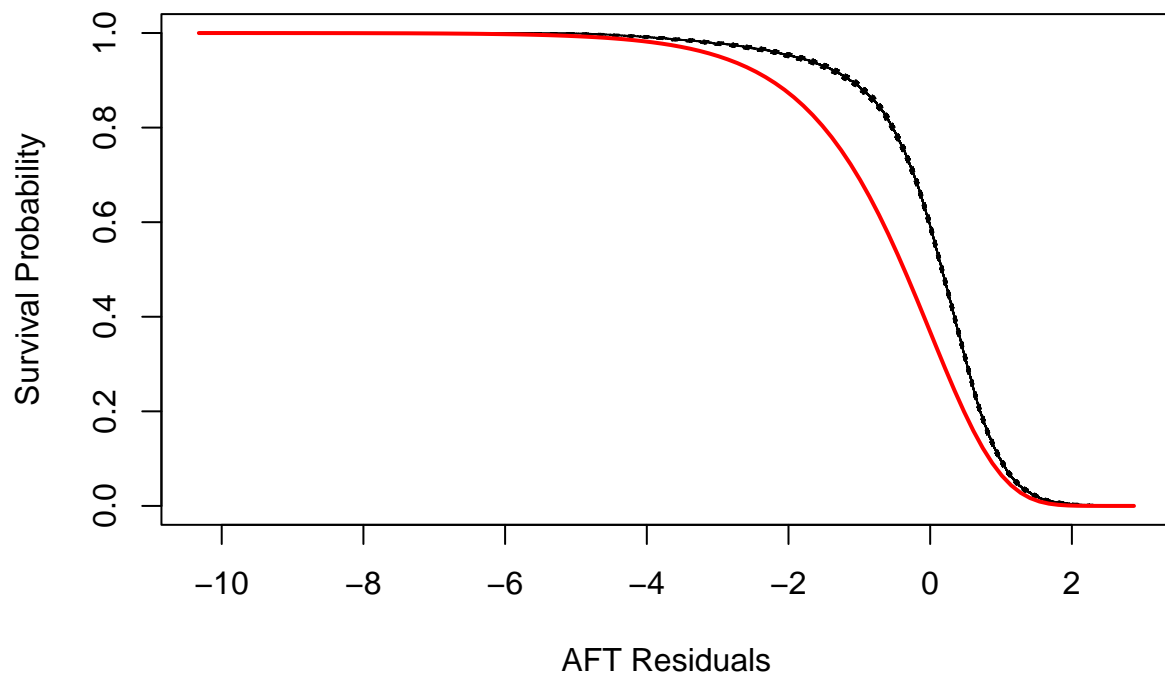
Table 4: AIC and BIC for Log-normal AFT Models by Race Encoding Method

	Race complete	Race Black/non-Black	Race white/non-white
AIC	301789.7	301786.8	301782.9
BIC	302702.1	302667.2	302663.3

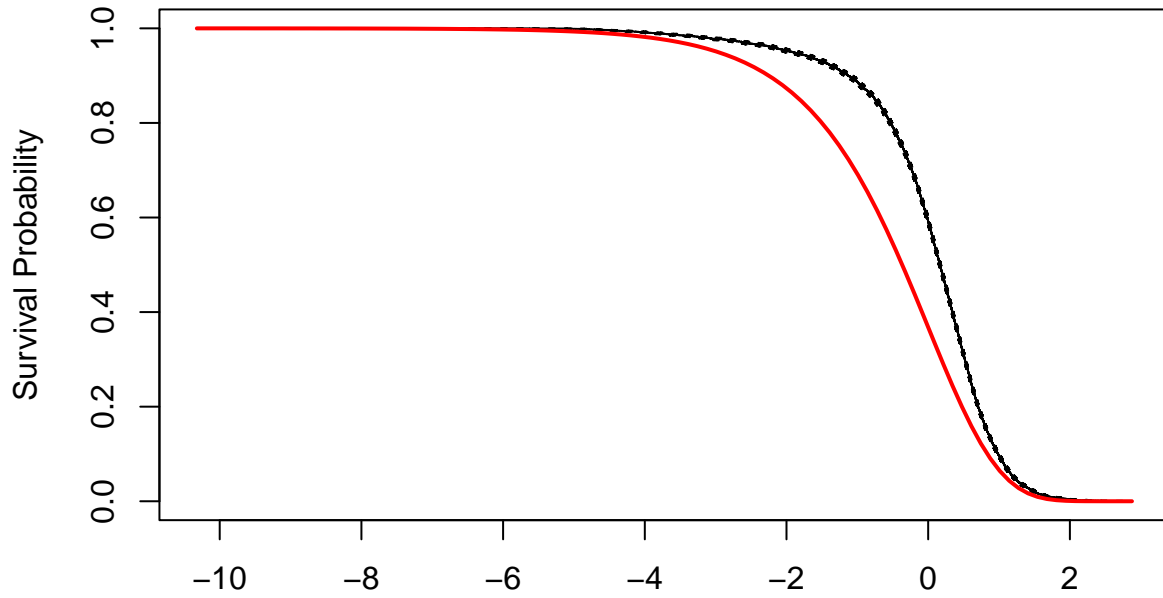
Log-Normal Model with All Confounders



Exponential Model with All Confounders



Weibull Model with All Confounders



Logistic Model with All Confounders

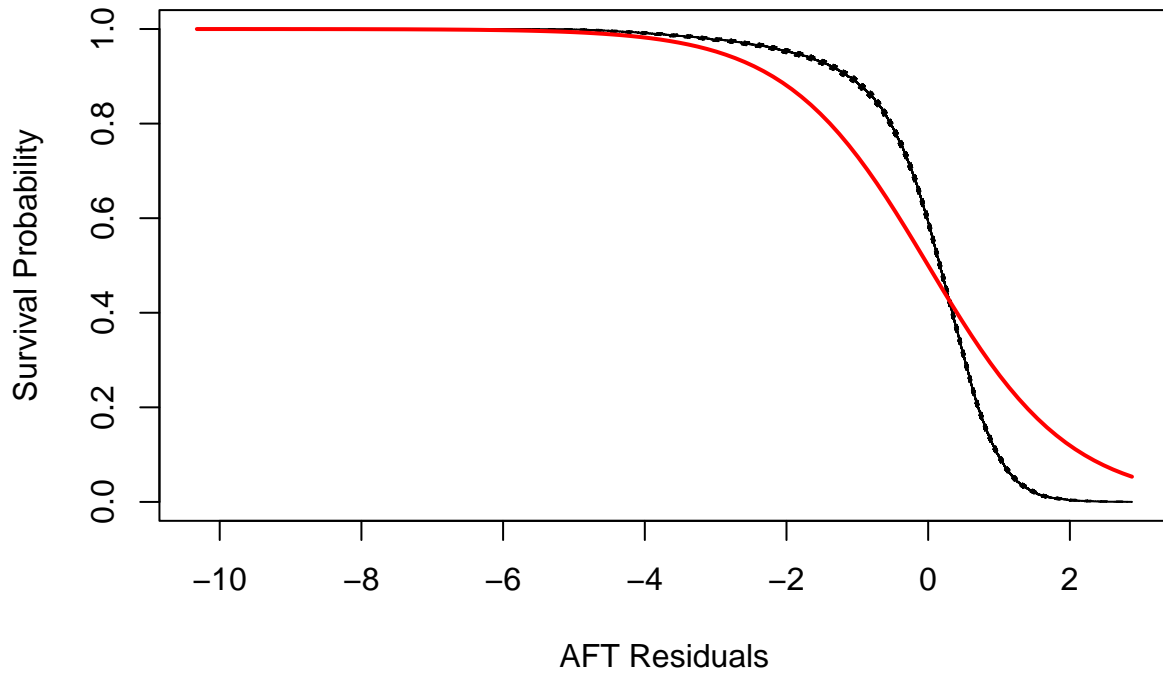


Table 5: AIC and BIC for Log-normal AFT Models by Race Encoding Method

	Race complete	Race Black/non-Black	Race white/non-white
AIC	144444.9	144443.1	144441.5
BIC	145298.2	145274.7	145273.2

- See Appendix for model assumptions, multicollinearity analysis, discussion of residuals
- Description of coefficients of interest, any significant results, and effect size.