An Analysis of Distress Tolerance as a Predictor of Early Treatment Dropout in a Residential Substance Abuse Treatment Facility

Riley Harper
University of North Carolina at Chapel Hill
December 12, 2023

1. Introduction

In the following pages, I will delve into the intricacies of an original study in the field of substance abuse treatment, offering a detailed analysis and reflection on the methodologies employed. As a fulfillment of the STOR 496 project requirements, this analysis covers the statistical tests used, their purposes, discusses the relevant R packages and functions, and interprets the outcomes. My primary goal is to showcase the knowledge and skills I've acquired in discrete event history analysis, through the lens of Dr. Daughters' research.

2. Original Study Synopsis

2.1 Research Question

The original study [Daughters et al., 2005] aimed to explore the complex relationship between distress tolerance and early dropout rates in residential substance abuse treatment programs. It focused primarily on assessing how distress tolerance predicts early treatment dropout, and secondarily on understanding the disparity between the influences of psychological and physical distress tolerance on these rates. This investigation was anticipated to yield valuable insights, informing the development of treatment strategies to enhance their effectiveness and ultimately reduce premature treatment dropout.

2.2 Patient Demographics

This study involved 122 individuals enrolled at the Salvation Army Harbor Light residential substance abuse treatment center in Northeast Washington, DC. The average age of participants was 40.3 years, with a majority of 70.5% being male and 95.1% identifying as African American. Educational backgrounds varied: 27.0% had less than a high school diploma, 43.4% had completed high school or equivalent, 20.5% had some college or technical education, and 9.1% held a college degree or higher. Substance use patterns showed 60.7% used crack/cocaine, 41.0% alcohol, 27.9% heroin, and 27.0% cannabis on a weekly basis in the past year. Treatment durations signed for by the participants ranged from 30 days (45.1% of participants) to 180 days (25.4%).

2.3 Recruitment and Screening Process

Out of 144 candidates initially approached within their first week at the center, 16 declined participation. Six more were excluded due to psychosis, determined using the Structured Clinical Interview for DSM–IV (SCID-IV) [Kübler, 2013]. The time from arrival at the facility to study participation averaged 2.6 days. All participants had to demonstrate drug abstinence and complete detoxification before joining the treatment center.

2.4 Study Procedures and Instruments

The study combined the SCID-IV with a series of seven self-report questionnaires and four distress tolerance tasks, with varied order across participants. Performance in the tasks influenced compensation, ranging from $5 to $15. Key tools included the Positive and Negative Affect Schedule (PANAS) [Tran, 2013] for mood assessment, the Center for Epidemiological Studies–Depression Scale (CES-D) [Herge et al., 2013] for depressive symptoms, the Multidimensional Personality Questionnaire–Stress Reaction Subscale (MPQ-SR) [Patrick and Kramer, 2017] for stress-related traits,

and the Eysenck Impulsiveness Scale [Huang, 2022] for impulsivity levels. All showed acceptable to high reliability (PANAS: .89, CES-D: .76, MPQ-SR: .84, Eysenck: .76). Additionally, a polydrug use questionnaire assessed drug use across 10 categories. The Interpersonal Support Evaluation List (ISEL) evaluated perceived functional support (reliability: .74), and the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) measured readiness to change in substance use behavior (reliability: .69, .85, and .90 for different scales).

2.5 Distress Tolerance Tasks

Distress Tolerance Tasks are pivotal in the empirical investigation of psychological resilience and coping mechanisms. These tasks provide a standardized method to quantifiably measure how individuals withstand and respond to stressful or challenging situations. The empirical data gathered from such tasks are invaluable, as they offer objective insights into human behavior under stress. This, in turn, aids in the development of targeted interventions and therapeutic strategies. Overall, the results we derive from Distress Tolerance Tasks contribute significantly to our understanding of various psychological conditions, particularly those related to anxiety and stress disorders. By employing these tasks, researchers can establish a more robust foundation for psychological theories and practices, ensuring they are grounded in observable, measurable phenomena.

2.5.1 Psychological Stressor 1: Paced Auditory Serial Addition Task (PASAT)

This task [Tombaugh, 2006] involved participants adding numbers that were sequentially displayed on a computer screen, with each new number requiring addition to the one prior. The task was structured in three levels of increasing difficulty, characterized by shorter intervals between number presentations. Participants were informed they could opt out of the final level at any time, with their earnings from the session being contingent on their task performance. Unknown to the participants, the task was designed to end after 7 minutes. Distress tolerance was gauged by the latency in seconds until task termination. Additionally, dysphoria was measured using a four-item scale assessing anxiety, difficulty concentrating, irritability, and frustration, showing acceptable reliability (0.69). To evaluate the increase in psychological stress, dysphoria levels were recorded both at the session's start and after the second PASAT level, thus avoiding confounds with termination latency.

2.5.2 Psychological Stressor 2: Computerized Mirror-Tracing Persistence Task (MTPT-C)

This task [Brown et al., 2018] required participants to trace a red dot along a star using a computer mouse, which was programmed to move the dot in the opposite intended direction. Deviations from the path or a stall of more than 2 seconds triggered a loud buzzer, resetting the dot to the starting position. Participants were allowed to terminate the task at any point, with their performance impacting their financial compensation. The task was secretly set to conclude after 5 minutes. Distress tolerance was again measured by the time taken to terminate the task. Additionally, the number of errors per second was recorded to control for skill level in persistence. Since the MTPT only involved a single level, dysphoria could not be assessed without confounding termination latency.

2.5.3 Physical Stressor 1: Breath Holding (BH)

This task, previously shown to predict cessation attempts in smokers [Brown et al., 2002], was used as a physical challenge. Participants were instructed to hold their breath for as long as possible while an experimenter timed the duration using a stopwatch. Persistence was quantified as the time in seconds before taking a breath.

2.5.4 Physical Stressor 2: Cold Pressor Task (CPT):

This task [von Baeyer et al., 2005] involves immersing the participant's nondominant hand and forearm in ice water

(33° Fahrenheit; SD 1°) up to a marked point. Participants were instructed to keep their hand submerged for as long as they could, with the option to remove it at any time. Unbeknownst to them, the task was set to end after 5 minutes. Persistence was measured by the duration before removing the hand from the water.

2.6 Group Status

Participants were grouped based on the duration of their treatment adherence, irrespective of the initially planned length of stay. Early dropouts, comprising 20 individuals, were those who did not complete the initial 30-day treatment period. This group included participants who, against medical advice, chose to voluntarily exit the program, numbering eight, and those dismissed for substance use during the treatment, totaling twelve. In contrast, completers were defined as individuals who successfully met or exceeded the 30-day treatment threshold. This group consisted of 102 participants who adhered to the program's minimum duration requirement, reflecting a commitment to the prescribed treatment course. A further discussion of results will be carried on in the results section of this paper.

3. Methods

3.1 Outline

My analysis aimed to replicate and extend the methodologies used in the original paper to deepen my understanding of the employed techniques. Initially, I focused on recreating the descriptive statistics presented in the study, including the mean and standard deviation calculations for key variables. Following this, I conducted correlation tests to examine linear relationships among the covariates. To further explore these relationships, I performed Welch's Two Sample t-tests to assess the significance of group mean differences. The final step involved implementing a stepwise approach in a Cox Proportional Hazards model, which allowed me to identify the most significant factors among the four psychological and physical tasks discussed in the original paper.

3.2 Descriptive Statistics

First, we will look into the use of descriptive statistics [Anderson et al., 2023], namely the mean and standard deviation. Descriptive statistics are an essential tool in statistics, providing a simple summary about the sample at hand. They do this by simplifying large amounts of data into measures of central tendency and variability. Measures of central tendency describe the center point of a data set–the mean$_{(1)}$, median$_{(2)}$, and mode$_{(3)}$. The mean provides the average value, the median denotes the middle value, and the mode represents the most frequently occurring value. In general, for a set of data, X, with n elements ordered from least to greatest,

$$\textbf{Mean} = \bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1}$$

$$\textbf{Median} = \begin{cases} x_{\frac{n+1}{2}}, & \text{if } n \text{ is odd} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{if } n \text{ is even} \end{cases} \tag{2}$$

$$\textbf{Mode} = \arg\max_{x_i} P(X = x_i) \tag{3}$$

The other central components of variability describe the spread or dispersion of the data through the range$_{(4)}$, variance$_{(5)}$, and standard deviation$_{(6)}$. The range is the difference between the highest and lowest values. Variance and

standard deviation are measures of high far individual data points deviate from the mean. When computing the variance and standard deviation of a sample rather than a population $n-1$ is used rather than n. This is known as Bessel's correction and corrects for the bias in the estimation of the population variance and standard deviation.

$$\textbf{Range} = \max(x_i) - \min(x_i) \tag{4}$$

$$\textbf{Variance} = s^2 = \begin{cases} \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2, & \text{if each observation is equally likely} \\ \sum_{i=1}^{n} p_i \cdot \left(x_i - \sum_{j=1}^{n} p_j x_j\right)^2, & \text{if observations are not equally likely} \end{cases} \tag{5}$$

*p_i represents the probability of each observation

$$\textbf{Standard Deviation} = s = \begin{cases} \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2}, & \text{if each observation is equally likely} \\ \sqrt{\sum_{i=1}^{n} p_i \cdot \left(x_i - \sum_{j=1}^{n} p_j x_j\right)^2}, & \text{if observations are not equally likely} \end{cases} \tag{6}$$

*p_i represents the probability of each observation

The above variance and standard deviation are given for a sample and thus have been denoted with s rather than for the population, in which case they would have been denoted σ. These six measures used in descriptive statistics provide key insights into the central tendency and variability of your data.

3.3 Correlation Test

Understanding the relationship between two variables is often a key objective. One of the most common measures used to assess the strength and direction of a linear relationship between two variables is the sample correlation coefficient$_{(7)}$. This coefficient r_{XY}, typically denoted as r, for two random variables X and Y can be found as follows,

$$\textbf{Correlation Coefficient} = r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{7}$$

Often in software optimizing for computational efficiency requires the avoidance of computing the mean. The following rearrangement of the correlation coefficient$_{(8)}$ allows for this computation to be done without directly calculating the mean,

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \tag{8}$$

The Correlation Coefficient computed above is the most commonly used and is known as the Pearson Correlation Coef-

ficient (PCC). There have also been other correlation coefficients developed–such as Spearman's rank correlation–which are more robust methods and have the potential to demonstrate non linear correlation between variables.

3.4 Student's t-Test

The Student's t-test [The Editors of Encyclopaedia Britannica, 2023] is a fundamental tool for comparing the means of two groups. This test is instrumental in determining whether the observed differences between these groups are significant or the result of random variation. At it's core, the t-test operates under the principle of hypothesis testing. It helps decide whether data can convincingly show a significant difference between two sets of numbers, such as measurements, scores, or observations. The term "Student" refers to the pseudonym used by William Sealy Gosset, who developed the test.

The t-test assumes that the data follows the Student's t-distribution. This assumption is crucial when the scaling term, affecting the data's spread or dispersion, is unknown and must be estimated from the data itself. Several types of t-tests exist, each used in different situations, including the one-sample t-test, independent two-sample t-test, and paired sample t-test. A common application of the t-test is in experimental design, where it helps to determine if a treatment or condition has a statistically different effect from a control condition.

Each of the following tests involve the use of t-scores. A t-score table can be used to mark critical values for different levels of significance, α, and different degrees of freedom, DF. A t-score table is pictured in the appendix.

In the one sample t-test we're often looking to confirm the null hypothesis $H_0 : \mu = 0$ against the alternative hypothesis $H_a : \mu \neq 0$.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \tag{9}$$

*DF $= n - 1$

Alternatively, when given two groups or variables we may consider the independent two-sample t-test which assumes the two sample sizes are equal and that the two distributions have equal variance. This t-test is defined as,

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s_{\bar{x}}^2 + s_{\bar{y}}^2}} \tag{10}$$

In the case when there may be unequal sample sizes but the variances of the two distributions are similar such that $\frac{1}{2} < \frac{s_X}{s_Y} < 2$ we can use the pooled variance. When you have two independent samples with similar variances, the pooled variance, denoted as s_p^2, is used as an estimate of the common variance of the two populations. The pooled variance is calculated as follows:

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2} \tag{11}$$

*s_X^2, s_Y^2 are the sample variances of the two samples, and n_X, n_Y are the sample sizes

The t-statistic for testing the difference between two means, assuming similar variances but with unequal sample sizes, is given by:

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \tag{12}$$

*$\bar{x}, \bar{y}$ are the sample means of the two samples, and s_p is the pooled standard deviation–the square root of the pooled variance s_p^2

*DF $= n_X + n_Y - 2$

In the case when there may be unequal sample sizes and unequal variances such that $s_X > 2s_Y$ or $s_Y > 2s_X$ we can use the Welch's t-test [West, 2021]. For this test, we'll utilize $s_{\bar{\Delta}}$ to estimate the standard error of the difference between the two sample means. The formula for $s_{\bar{\Delta}}$ is given by,

$$s_{\bar{\Delta}} = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \tag{13}$$

Then, the Welch's t-test can be calculated as follows,

$$t = \frac{\bar{x} - \bar{y}}{s_{\bar{\Delta}}} \tag{14}$$

*DF can be calculated using the Welch-Satterthwaite equation

The Welch-Satterwaite equation is used to approximate degrees of freedom without an assumption that the underlying population variances, σ_i^2, are equal and is also known as the pooled degrees of freedom. This equation is given by,

$$\text{DF} \approx \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{\left(\frac{s_X^2}{n_X}\right)^2}{n_X - 1} + \frac{\left(\frac{s_Y^2}{n_Y}\right)^2}{n_Y - 1}} \tag{15}$$

3.5 Cox Proportional Hazards Model

The Cox Proportional Hazards model is a cornerstone of survival analysis, a field of statistics that focuses on the expected duration of time until an event of interest occurs. Introduced by Sir David Cox in 1972 [Cox, 1972], this model has since become one of the most widely used methods for analyzing survival data, particularly in the context of medical research and reliability engineering.

Unlike traditional regression models that predict a numeric value or classification, the Cox model is designed to understand and quantify the effect of various covariates on the hazard, or the instantaneous risk of experiencing the event at a given time, conditional on survival until that time. One of its most defining features is the assumption of proportional hazards–the ratios of the hazards for any two individuals are constant over time, regardless of the value of the survival time.

This section aims to delve into the fundamentals of the Cox Proportional Hazards model by discussing how the model is formulated, interpreted, and how it can be applied to real-world datasets to glean insights into factors that influence the time until an event occurs.

Before delving into the Cox Proportional Hazards model we will first define a much simpler model called the Kaplan

Meier Model which is the foundational model used in survival analysis. The Kaplan-Meier estimator is defined as,

$$\hat{S}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}) \tag{16}$$

*t_i are the times when at least one event has occurred

*d_i is the number of events at time t_i

*n_i is the number of individuals known to have survived up to time t_i and are still at risk

$\hat{S}$(t) can be interpreted as the estimated probability that an individual will remain event-free beyond a certain point–t. The Kaplan-Meier curve is thought of as a stepwise function since it decreases at each time point where an event occurs and is assumed to be constant between these points–it is a decreasing function. The estimator is also non-parametric such that it makes no assumption of the underlying distribution for survival times.

With an understanding of the Kaplan-Meier estimator in place, we now turn our attention to the Cox Proportional Hazards Model, a more advanced tool in survival analysis. While the Kaplan-Meier estimator is useful for estimating survival probabilities, it does not allow for the direct assessment of how specific factors or covariates might affect survival times. This is where the Cox Proportional Hazards Model becomes more useful.

The Cox Proportional Hazards model is a semi-parametric model used to investigate the effect of several variables on the survival time of subjects. One of the key features of this model is its ability to handle both categorical and continuous variables, providing a way to understand how different factors contribute to the hazard–the instantaneous risk of experiencing the event of interest. The model is given by the formula,

$$h(t) = h_0(t) \exp\left(\sum_{i=1}^{p} \beta_i x_i\right) = h_0(t) \cdot e^{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p} \tag{17}$$

*$h_0(t)$ is the baseline hazard which denotes the hazard rate for a baseline level of covariates

*$\beta_1, \beta_2, \cdots \beta_p$ are the coefficients of the covariates

*$x_1, x_2, \cdots x_p$ are the covariates

The baseline hazard quantifies the instantaneous risk of an event occurring at time t, assuming no influence from the covariates such that all covariates are zero or their baseline level. The baseline hazard is demonstrated below,

$$h(t) = h_0(t) \exp\left(\sum_{i=1}^{p} \beta_i(0)\right) = h_0(t) \exp(0) = h_0(t) \tag{18}$$

As previously stated the model is semi-parametic. This is due to the model not specifying a function for the baseline hazard and assuming a multiplicative effect from the covariates on the hazard rate.

3.6 Log Rank Test

The Log Rank Test [Goel et al., 2010] is a non-parametric test used in survival analysis for comparing the survival distributions of two or more groups. The primary application of the Log Rank Test is to assess whether there are statistically significant differences between the survival curves of different groups, often in the context of clinical trials. This test is based on the survival probabilities estimated at various time points, typically when an event (like death, failure, or relapse) occurs.

The key principle behind the Log Rank Test is to compare the observed number of events in each group at each time point with the expected number of events, assuming no difference between the groups. The test statistic is then derived from the sum of these comparisons over all time points.

The formula for the Log Rank Test statistic is as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_{1i} - E_{1i})^2}{E_{1i}} + \frac{(O_{2i} - E_{2i})^2}{E_{2i}} \tag{19}$$

*O_{ji} is the observed number of events in group j at time i,

*E_{ji} is the expected number of events in group j at time i,

*k is the number of time points.

In practice, the Log Rank Test involves constructing a table that captures the number of subjects at risk, the number of events, and the expected events under the null hypothesis for each time point. Next, the test calculates the total observed and expected events across all time points for each group. The test statistic is compared against a chi-square distribution to determine the p-value, which indicates the probability of observing such a difference if the null hypothesis–which states that there is no difference in survival between groups–were true.

An assumption of the Log Rank Test is that of proportional hazards, implying that the ratio of the hazard functions of the groups being compared is constant over time. However, this assumption does not imply that the hazard functions themselves are constant over time.

The Log Rank Test is favored for its simplicity and the minimal assumptions it makes about the data, and it is particularly powerful in large sample sizes or when the event of interest does not occur often. However, it is less sensitive when the hazard rates cross over time, which may occur in some practical scenarios. In this case, another test which can be used is the Wilcoxon-Breslow-Gehan test. Overall, the Log Rank Test offers a straightforward method to compare survival times across different groups.

3.7 Checking Linearity

Checking linearity is also a critical step in survival analysis, especially when using models like the Cox Proportional Hazards model, which assumes a linear relationship between the log of the hazard function and the covariates. This assumption of linearity influences the interpretation of the model's coefficients along with the validity of its predictions. Verifying this assumption is important to ensure the accuracy and reliability of the model's outcomes.

One common method for checking linearity in the Cox model is through examining Martingale residuals [Gillespie, 2006]. Martingale residuals are the differences between the observed and expected numbers of events, given the covariates for each individual. By plotting these residuals against each of the covariates, one can visually inspect for any systematic patterns that would indicate non-linearity. If the relationship is linear, the plot should show a random scatter of points around zero without any discernible pattern.

Martingale Residuals can be calculated as,

$$\textbf{Martingale Residual} = \delta_i - \hat{H}_i(t) \tag{20}$$

*δ_i is the event indicator (1 if the event occurred, 0 otherwise)

*$\hat{H}_i(t)$ is the cumulative hazard function estimated at the observed survival time for the ith individual

The equation for the estimated cumulative hazard function [Bradburn et al., 2003], $\hat{H}_i(t)$, in the context of the Cox Proportional Hazards model involves integrating the hazard rates over time. This can be expressed as follows:

$$\hat{H}_i(t) = \int_0^t \hat{h}_i(u)\, du \tag{21}$$

*$\hat{h}_i(u)$ represents the estimated hazard function for the ith individual at a given time u

*u ranges from the start of observation at 0 up to time t

*$\int_0^t$ indicates that we are summing up the hazard rates over the interval from 0 to t

This cumulative hazard function is not directly computed as a closed-form integral because the Cox model does not specify the baseline hazard function $h_0(t)$ explicitly. Instead, it is estimated using the data and the proportional hazards assumption.

To further test linearity another approach involves adding interaction terms between covariates and a function of time, such as log(time), to the Cox model. If these interaction terms are statistically significant, this may indicate a violation of the linearity assumption.

In cases where linearity is violated, transforming covariates to instead use the log or square root transformations is a viable approach to achieve linearity. These transformations can be particularly useful with skewed data or when the relationship between the covariates and the log hazard is not linear. After transforming the covariates, it is important to recheck the model to ensure that the transformed variables now exhibit a linear relationship with the log hazard.

Overall, checking linearity in survival analysis involves using residual plots, adding interaction terms, applying transformations to covariates, and reevaluating the model. This process helps validate the assumptions of the Cox Proportional Hazards model and ensures its interpretations and predictions are accurate.

3.8 Checking Proportional Hazards Assumption

The proportional hazards assumption is a key aspect of the Cox Proportional Hazards model. This assumption states that the hazard ratios for different levels of a covariate will be constant over time. Violation of this assumption can lead to biased estimates and incorrect conclusions so several techniques can be employed to evaluate its validity.

A common approach to checking the proportional hazards assumption is to use the Schoenfeld residuals. These residuals represent the difference between observed and expected covariate values at each event time. By plotting the Schoenfeld residuals against time, one can assess if the effect of the covariates changes over time, which would violate the proportional hazards assumption. A pattern which is not random in these plots suggests that the proportional hazards assumption may not hold.

Schoenfeld Residuals for covariate j can be calculated as:

$$\textbf{Schoenfeld Residual for Covariate j} = (x_{ij} - \bar{x}_j(t))(\delta_i - \hat{h}_i(t)) \tag{22}$$

*x_{ij} is the value of the jth covariate for the ith individual

*$\bar{x}_j(t)$ is the average value of the jth covariate for individuals at risk at time t

*δ_i is the event indicator (1 if the event occurred, 0 otherwise)

*$\hat{h}_i(t)$ is the estimated hazard at time t for the ith individual

Additionally, a visual inspection of survival plots for various subgroups within the study can also offer valuable insights on this assumption. This method entails comparing the survival
curves of different groups over time. If these curves maintain a consistent relationship with each other — if they appear to be parallel when plotted on top of one another — it suggests that the proportional hazards assumption may be valid. This is true since parallel survival curves imply that the relative risk or hazard ratios between the groups do not change significantly over the study period, which aligns with the assumption that these ratios are constant over time.

A third method for checking the proportional hazards assumption involves incorporating time-dependent covariates in the Cox model. An interaction between covariates and time that significantly improves model fit could signal a breach of the assumption since the hazards would then clearly not be proportional across the duration of the study.

In summary, validating the proportional hazards assumption involves a multifaceted approach encompassing Schoenfeld residuals analysis, graphical methods, time-dependent covariate inclusion, and formal statistical testing. Ensuring this assumption holds is crucial for the reliability and accuracy of model interpretations.

4. Results

Each of the methods we discussed in the previous section will be utilized in this results section, for more information on the methods you should consult the preceding section. Following the comprehensive evaluation of the Cox Proportional Hazards model's assumptions, we now present our results. The work on the data begins with cleaning. This was not a difficult task for this project but did involve removing several rows containing either all NAs or several NAs. It is noted in the original paper that the rows which I chose to remove, containing only several NAs, were kept for measuring some tasks but not the tasks which contained an NA. Since these individuals were measured differently than others, I felt it was best to either refrain from using the individuals, as I did, or to have imputed their data for the missing tasks. After cleaning, my data contained 22 early dropouts and 103 completers. Also notably, my data cleaning did not account for substance use frequency due to my lack of understanding the original decision boundary of the study. It was stated in the original study that there were 20 early dropouts and 102 completers used in their analysis. Next, I looked into the average time on the PASAT which was 217.31 seconds with a standard deviation of 166.79 seconds. It was noted in the original paper that the average time on the PASAT was 208.7 seconds with a standard deviation of 165.2 seconds.

Next, I chose to use a Welch two sample t-test to compare the PASAT total time between individuals who dropped out before 30 days of treatment and those who completed the treatment. The test did not find a statistically significant difference in the average PASAT Total Time between the dropout group (mean = 173.36 seconds) and the completer group (mean = 226.7 seconds), t(31.33) = -1.3957, p = 0.1726. It's important to note that the original paper did not use a Welch's two sample t test since it was making a comparison with dysphoria which was not data I was provided. Importantly, the test which I have performed, while significant, was not the same as was performed in the original paper. This lead to different formulas for calculating the degrees of freedom, as was noted in my methods section.

After this, I looked into the other psychological distress task–MTPT-and found the average to be 197.62 seconds with a standard deviation of 94.84 seconds. In the original paper these were found to be 197.1 seconds and 95.9 seconds respectively, however, a correlation test between the PASAT and MTPT measures of psychological distress re-

sulted in a similar conclusion of 0.35 compared to the original paper's 0.38-both with $p < 0.001$. The last descriptive statistics were for BH with an average of 30.13 seconds and a standard deviation of 13.55 seconds and for CPT with an average of 97.26 seconds and a standard deviation of 102.86 seconds. In the paper these were reported as 30.12 seconds, 13.8 seconds, 99.97 seconds, and 104.6 seconds respectively. Running the same correlation test between these two measures of physical distress tolerance as I have previously we obtain a correlation coefficient of 0.26 with $p < 0.01$. For each of these correlation coefficients, I also computed their 95% which came out to 0.18 to 0.49 and 0.09 to 0.42, respectively. From here, I partitioned the physical and psychological measure of distress into a group containing only the early dropouts and a group containing only the completers. I then once again calculated the mean and standard deviation for both and will report these compared side by side in the figures below.

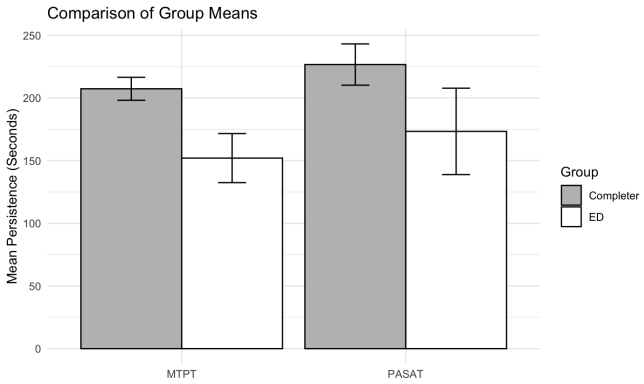

Figure 1: Psychological Distress Measures

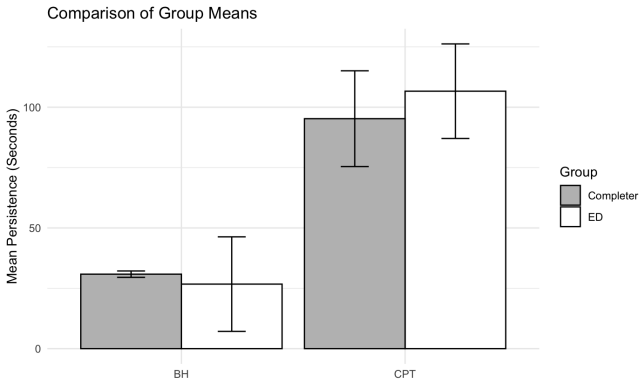

Figure 2: Physical Distress Measures

Notably, there were no group differences between the two physical distress measures while there does appear to be $\tilde{5}0$ second difference in the group means between the completer and early dropout groups for the two psychological distress measures. This was also found in the paper, although our numbers slightly differ likely due to different sample sizes as noted in the beginning of results. From here, our analysis proceeded with the application of Cox Proportional Hazards (Cox-PH) models to investigate the relationship between various distress measures on the duration of treatment. I selected three models to consider. The first model incorporated all four distress measures–Breath Holding Duration, Cold Pressor Duration, PASAT Total Time, and Mirror Tracing Total Time. The second model focused on the psychological distress measures-PASAT Total Time and Mirror Tracing Total Time. The third model examined the physical distress measures-Breath Holding Duration and Cold Pressor Duration. Each model was designed to elucidate the distinct and combined impacts of these measures on the treatment duration. In R, the models were written as follows,

```
# Model 1: All Measures
cox.mod <- coxph(Surv(‘Days in Treatment‘, drop30d) ~
    ‘Breath Holding Duration‘ + ‘Cold Pressor Duration‘ +
    ‘PASAT Total Time‘ + ‘Mirror Tracing Total Time‘, data = datamod)


# Model 2: Psychological Distress Measures
cox.mod2 <- coxph(Surv(‘Days in Treatment‘, drop30d) ~
```

```
    `PASAT Total Time` + `Mirror Tracing Total Time`, data = datamod)

# Model 3: Physical Distress Measures
cox.mod3 <- coxph(Surv(`Days in Treatment`, drop30d) ~
    `Breath Holding Duration` + `Cold Pressor Duration`, data = datamod)
```

The coxph() function from the survival package in R was used to generate the Cox Proportional Hazards models, and the summary() function in R was utilized to extract the coefficients, standard errors, z-scores, and p-values for each predictor in the models. It is of note, that datamod is our original data set after it has been cleaned and contains both the early dropouts and completers.

From the summary of the first model, we see that the most extreme association is a slightly negative association of Breath Holding Duration with Days in Treatment with coef = -0.024, p = 0.17. Following this, Cold Pressor Duration and PASAT Total Time exhibited non-significant ($p > 0.05$) associations which were both relatively small. Finally, the Mirror Tracing Total Time was found to be significant with coef = -0.005, p = 0.04.

From the summary of the second model, we see that both PASAT Total Time and Mirror Tracing Total Time showed negative associations with the outcome, with MTPT being significant coef = -0.00519, p = 0.03.

From the summary of the third model, we see that both neither 'Breath Holding Duration' nor 'Cold Pressor Duration' showed significant associations with Days in Treatment.

The results from these Cox Proportional Hazards models provide valuable insights into how different distress measures impact expected treatment duration. The significant association of Mirror Tracing Total Time in both the full model and the psychological distress-only model suggests its relevance in predicting treatment outcomes, while the lack of significant findings for physical distress measures might indicate a lesser impact of these measures on treatment duration, or alternatively, the need for a larger sample size with different physical measures to detect an effect.

Next, residual analyses to test the linearity assumption on the second model were conducted using Martingale and Deviance residuals. Below martingale residuals were plotted against the fitted values from Model 2 and a smoothing line was added to aid in the identification of any non-linear patterns. The R code for generating this plot is as follows,

```
plot(predict(cox.mod2), residuals(cox.mod2, type = "martingale"),
    xlab = "fitted values", ylab = "martingale residuals",
     main = "Residual Plot", las = 1)
abline(h = 0)
lines(smooth.spline(predict(cox.mod2),
      residuals(cox.mod2, type = "martingale")), col = "red")
```

Here, a lack of a pattern or trendline in the plot suggests that the linearity assumption is reasonable for the variables in the model.

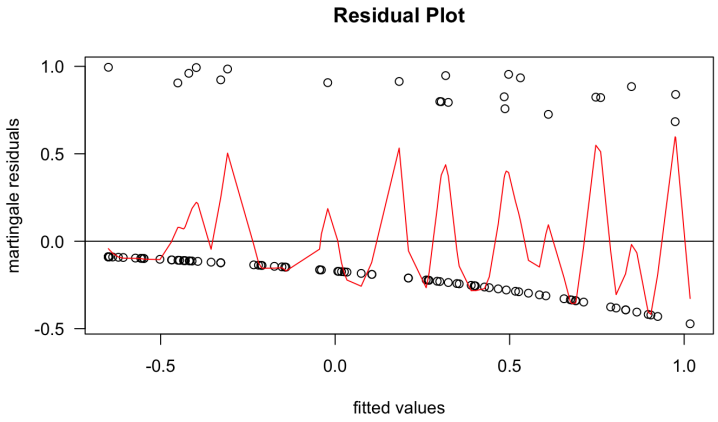

Figure 3: Martingale Residual Plot

Since the red line above clearly lacks any non-linear patterns this plot supports the linearity assumption for the psychological distress measures which were used in the model. Next, deviance residuals were also examined to ensure the robustness of the linearity assumption. Similar to the Martingale residuals, a plot of Deviance residuals against fitted values was constructed, looking for patterns that might indicate violations of the linearity assumption. The R code for generating this plot is as follows,

```
plot(predict(cox.mod2), residuals(cox.mod2, type = "deviance"),
    xlab = "fitted values", ylab = "deviance residuals",
    main = "Residual Plot", las = 1)
abline(h = 0)
lines(smooth.spline(predict(cox.mod2),
    residuals(cox.mod2, type = "deviance")), col = "red")
```

No patterns were identified as the residuals appear to be randomly scattered around the horizontal line at zero, further confirming the linearity assumption for the data.

The final assumption to look into was the proportional hazards assumption which was evaluated using the Schoenfeld residuals. The cox.zph() function from the survival package in R was utilized to perform this assessment.

The Schoenfeld residuals test is used to determine if the effects of predictors are constant over time. A global test and individual tests for each covariate were conducted. The R code used for these tests is as follows:

```
# Global test for the model
cox.zph(cox.mod)
```

The output of this test includes a global chi-squared statistic and individual statistics for each predictor. A non-significant global test suggests that the assumption of proportional hazards is reasonable for the model. In our case the global chi-squared statistics was 5.7, $p = 0.06$ which is non-significant.

To further test this, individual plots for each covariate in the second model were generated to visually assess the proportional hazards assumption. The R code is as follows,

```
# Time-varying effect of PASAT Total Time
plot(cox.zph(cox.mod2)[1], xlab = "Transformed Time",
     ylab = "Scaled Schoenfeld Residuals",
     main = "Time-Varying Effect of PASAT Total Time")

# Time-varying effect of Mirror Tracing Total Time
plot(cox.zph(cox.mod2)[2], xlab = "Transformed Time",
     ylab = "Scaled Schoenfeld Residuals",
     main = "Time-Varying Effect of Mirror Tracing Total Time")
```

The plots of the scaled Schoenfeld residuals should show no significant patterns or trends over time if the assumption holds.

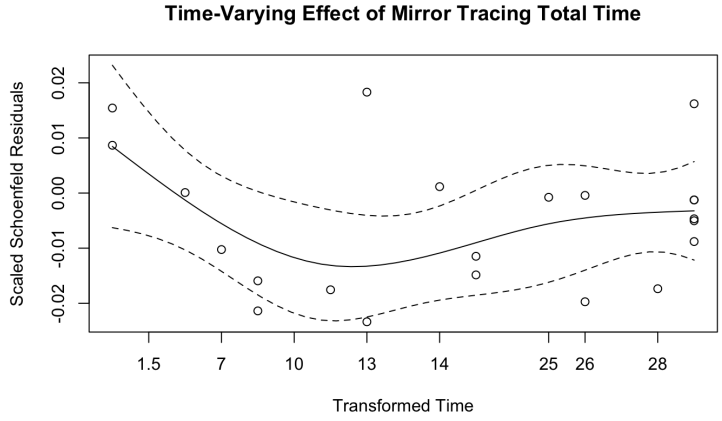

Figure 4

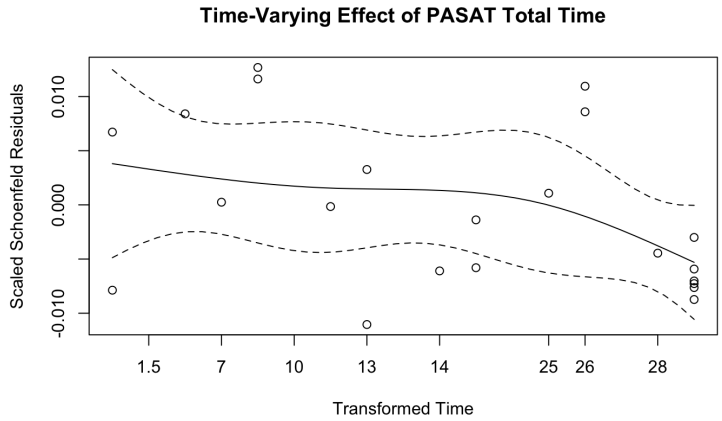

Figure 5

Departures from the horizontal line at $y = 0$ can be indicative of non-proportional hazards, as the assumption of proportional hazards posits that estimates for each covariate should remain relatively constant over time. In the analysis

of our second model, there were no significant deviations from this line for either covariate. It is important to consider the scale on the y-axis, which in this case is quite small. This suggests that any minor departures from the horizontal line are not substantial. Therefore, the tests we've performed for the proportional hazards assumption indicated that the constant hazard ratios for the predictors over time was a reasonable assumption in our models. This supports the use of the Cox Proportional Hazards models for the data and the interpretations we've derived from them.

5. Analysis

This analysis report has demonstrated the utility of Cox Proportional Hazards models in assessing the impact of psychological and physical distress measures on treatment duration. Consistent with the original paper, our findings underscore the more pronounced association of psychological distress measures—particularly the Mirror Tracing Persistency Task (MTPT)—with treatment dropout. A notable contribution of this analysis report is the rigorous testing of the Cox Proportional Hazards model assumptions, which were presumed but not empirically tested in the original work.

Future research on the original paper should consider employing more sophisticated statistical tests such as the Grambsch and Therneau test, which provides a numerical method for analyzing the correlation of scaled Schoenfeld residuals. Moreover, the utilization of the Breslow estimator for handling tied events or expanding the sample size would likely reduce standard deviations and yield a more representative sample, potentially offering broader insights into the replicability of this research. Investigating a variety of physical and psychological distress measures would also be interesting to demonstrate whether the observed results are an artifact of the specific measures of distress tolerance employed or indicative of underlying psychological processes.

Overall, the both the original study and this analysis clearly demonstrate the importance of the Cox Proportional Hazards model and its significance in statistical methodologies, particularly in the context of psychological research.

6. Reflections

My journey through this research project has been as much about statistical understanding as it has been a journey into the intricate world of psychology and addiction. Although my statistical foundation was laid in prior coursework, my understanding of psychology, particularly regarding addiction, was increased substantially through interdisciplinary learning. The seminal work *Unbroken Brain* [Szalavitz, 2016], participation in BRANE lab weekly meetings and lab work, and an extensive repository of shared knowledge through Sharepoint provided a rich education for this study. I also began my dive into discrete event history analysis through this project and have read countless papers on the topic throughout the course of the semester.

This analysis report illuminated the relevance of survival analysis within original research that has garnered significant academic attention, being cited in nearly 400 subsequent papers. Initially, I harbored skepticism about the practical application of the Cox Proportional Hazards model. However, through hands-on application, my appreciation for the semi-parametric nature of the Cox model has grown, especially its flexibility in not requiring a predetermined baseline hazard function.

While it is common in psychology papers to postulate the broader implications of empirical findings, I found it both enlightening and rewarding to delve deeper into the data. Conducting additional tests and analyses beyond which were assumed to be true in the original paper not only strengthened my statistical abilities but also served to bolster the arguments presented by the original authors.

Appendix

	PROPORTION (a) IN ONE TAIL								
	.25	.20	.15	.10	.05	.025	.01	.005	.0005
	PROPORTION (a) IN TWO TAILS COMBINED								
df	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.578
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.600
3	0.765	1.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	1.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.689
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.660
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.290

Figure 6: T-Score Table [Cote et al., 2021]

References in the Discussion

[Anderson et al., 2023] Anderson, D., Sweeney, D. J., and Williams, T. A. (2023). Statistics. In *Encyclopedia Britannica*. Encyclopedia Britannica, Inc.

[Bradburn et al., 2003] Bradburn, M., Clark, T., Love, S., and Altman, D. (2003). Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. *Br J Cancer*, 89(3):431–436.

[Brown et al., 2002] Brown, R., Lejuez, C., Kahler, C., and Strong, D. (2002). Distress Tolerance and Duration of Past Smoking Cessation Attempts. *Journal of Abnormal Psychology*, 111:180–5.

[Brown et al., 2018] Brown, R., Overstreet, C., Sheerin, C., Berenz, E., Hawn, S., Pickett, T., McDonald, S., Danielson, C., and Amstadter, A. (2018). The Nomological Network of a Behavioral Distress Tolerance Task in Veterans. *J Trauma Stress*, 31(6):876–885.

[Clark et al., 2003] Clark, T., Bradburn, M., Love, S., and Altman, D. (2003). Survival analysis part I: basic concepts and first analyses. *Br J Cancer*, 89(2):232–238.

[Cote et al., 2021] Cote, L. R., Gordon, R. G., Randell, C. E., Schmitt, J., and Marvin, H. (2021). *Introduction to Statistics in the Psychological Sciences*. University of Missouri–St. Louis.

[Cox, 1972] Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

[Daughters et al., 2005] Daughters, S. B., Lejuez, C. W., Bornovalova, M. A., Kahler, C. W., Strong, D. R., and Brown, R. A. (2005). Distress tolerance as a predictor of early treatment dropout in a residential substance abuse treatment facility. *J Abnorm Psychol*, 114(4):729–734.

[Ekman, 2017] Ekman, A. (2017). Variable selection for the Cox proportional hazards model: A simulation study comparing the stepwise, lasso and bootstrap approach. Unpublished master's thesis.

[Gillespie, 2006] Gillespie, B. (2006). Checking Assumptions in the Cox Proportional Hazards Regression Model. In *Midwest SAS Users Group (MWSUG) Conference*, Dearborn, Michigan. University of Michigan.

[Goel et al., 2010] Goel, M. K., Khanna, P., and Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, 1(4):274–278.

[Herge et al., 2013] Herge, W., Landoll, R., and La Greca, A. (2013). Center for Epidemiologic Studies Depression Scale (CES-D). In Gellman, M. and Turner, J., editors, *Encyclopedia of Behavioral Medicine*. Springer, New York, NY.

[Huang, 2022] Huang, H. (2022). Impulsivity. *Encyclopedia MDPI*.

[Kübler, 2013] Kübler, U. (2013). Structured Clinical Interview for DSM-IV (SCID). In Gellman, M. and Turner, J., editors, *Encyclopedia of Behavioral Medicine*. Springer, New York, NY.

[Patrick and Kramer, 2017] Patrick, C. and Kramer, M. (2017). Multidimensional Personality Questionnaire (MPQ). In Zeigler-Hill, V. and Shackelford, T., editors, *Encyclopedia of Personality and Individual Differences*. Springer, Cham.

[Stewart, 2023] Stewart, K. (2023). Pearson's correlation coefficient. In *Encyclopedia Britannica*. Encyclopedia Britannica, Inc.

[Szalavitz, 2016] Szalavitz, M. (2016). *Unbroken Brain: A Revolutionary New Way of Understanding Addiction*. St. Martin's Press, New York.

[The Editors of Encyclopaedia Britannica, 2023] The Editors of Encyclopaedia Britannica (2023). Student's t-test. Encyclopedia Britannica.

[Tombaugh, 2006] Tombaugh, T. N. (2006). A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology*, 21(1):53–76.

[Tran, 2013] Tran, V. (2013). Positive Affect Negative Affect Scale (PANAS). In Gellman, M. and Turner, J., editors, *Encyclopedia of Behavioral Medicine*. Springer, New York, NY.

[von Baeyer et al., 2005] von Baeyer, C., Piira, T., Chambers, C., Trapanotto, M., and Zeltzer, L. (2005). Guidelines for the cold pressor task as an experimental pain stimulus for use with children. *J Pain*, 6(4):218–27.

[West, 2021] West, R. (2021). Best practice in statistics: Use the Welch t-test when testing the difference between two groups. *Annals of Clinical Biochemistry*, 58(4):267–269.