Does TDT Affect Prognosis of Acute Myeloid Leukemia (AML)

This document will serve as a hub for all the charts, tests, and conclusions made for this experiment based on our AML Data.

## Pre-processing

We loaded full-set data and check for missing data. Missing data was removed for some tests, but did not significantly affect outcome.

## Descriptive Statistics

This section will include the calculations that produces 3 tables for summary statistics:

1. Patient and treatment characteristics of all 10 patients and according to TDT groups.
2. Treatment outcomes CR/CRi, ED, and 2-year OS of all patients stratified for TDT groups and age less than or equal to 60 years vs > 60 years.
3. ORs for the achievement of CR/CRi, ED, HRs for OS according to the linear multivariable logistic regression models.

### Table 1: Patient and treatment characteristics

| Parameter | All patients | TDT (0-30 d) | TDT (31-60d) | TDT (>60d) |
| --- | --- | --- | --- | --- |
| Age at initial diagnosis (years) |  |  |  |  |
| Mean (SD) | 37.35(12.2) | 36.57(12.66) | 40.88(12.02) | 35.11(9.71) |
| Median (IQR) | 37(18.75) | 35(18.5) | 44(12.75) | 37(7) |
| Female sex, no./no.available (%) | 40/72 (55.56) | 22/47 (46.81) | 14/16 (87.5) | 4/9 (44.44) |
| ECOG status 0-1, no./no.available (%) | 72/72 (100) | 47/47 (100) | 16/16 (100) | 9/9 (100) |
| HCT-CI score 0-2, no./no.available (%) | 69/72 (95.83) | 46/47 (97.87) | 16/16 (100) | 7/9 (77.78) |
| ENL Risk 2022 Group, no./no.available (%) |  |  |  |  |
| Favorable | 3/72 (4.17) | 2/47 (4.26) | 0/16 (0) | 1/9 (11.11) |
| Intermediate | 27/72 (37.5) | 13/47 (27.66) | 9/16 (56.25) | 5/9 (55.56) |
| Adverse | 2/72 (2.78) | 2/47 (4.26) | 0/16 (0) | 0/9 (0) |
| Unknown | 40/72 (55.56) | 30/47 (63.83) | 7/16 (43.75) | 3/9 (33.33) |
| AML type, no./no. available (%) |  |  |  |  |
| De novo AML | 68/72 (94.44) | 44/47 (93.62) | 16/16 (100) | 8/9 (88.89) |
| sAML | 2/72 (2.78) | 2/47 (4.26) | 0/16 (0) | 0/9 (0) |
| unknown | 2/72 (2.78) | 1/47 (2.13) | 0/16 (0) | 1/9 (11.11) |
| Cytoreductive pretreatment, no./no. available (%) |  |  |  |  |
| Hydroxyurea | 29/72 (40.28) | 22/47 (46.81) | 4/16 (25) | 3/9 (33.33) |
| Cytarabine | 13/72 (18.06) | 10/47 (21.28) | 2/16 (12.5) | 1/9 (11.11) |
| None given | 41/72 (56.94) | 24/47 (51.06) | 11/16 (68.75) | 6/9 (66.67) |
| TDT, d |  |  |  |  |
| Mean (SD) | 33.56(35.41) | 16(9.03) | 42.5(9.49) | 42.5(9.49) |
| Median (IQR) | 26.5(27.5) | 15(17.5) | 42(17.75) | 42(17.75) |
| WBC, x10^9/L |  |  |  |  |
| Mean (SD) | 51.74(78.87) | 64.03(89.41) | 22.6(37.63) | 22.6(37.63) |
| Median (IQR) | 15.66(55.55) | 21.92(74.8) | 11.65(17.88) | 11.65(17.88) |
| LDH (U/L) |  |  |  |  |
| Mean (SD) | 726.23(792.2) | 841.05(932.39) | 461.19(287.11) | 461.19(287.11) |
| Median (IQR) | 487(601) | 573(610.25) | 334(258.5) | 334(258.5) |
| Bone marrow blasts (%) |  |  |  |  |
| Mean (SD) | 50.83(21.28) | 55.65(19.94) | 43.69(22.5) | 38.37(19.25) |
| Median (IQR) | 52.67(39.23) | 56.52(28.7) | 45.52(34.07) | 30(18.61) |
| Karyotype, no./no. available (%) |  |  |  |  |
| No growth | 38/72 (52.78) | 27/47 (57.45) | 9/16 (56.25) | 2/9 (22.22) |
| Normal | 25/72 (34.72) | 12/47 (25.53) | 7/16 (43.75) | 6/9 (66.67) |
| Not Done | 4/72 (5.56) | 3/47 (6.38) | 0/16 (0) | 1/9 (11.11) |
| Treatment regimen (%) |  |  |  |  |
| H7+3 | 70/72 (97.22) | 45/47 (95.74) | 16/16 (100) | 9/9 (100) |
| HIDAC +/- Doxo | 2/72 (2.78) | 2/47 (4.26) | 0/16 (0) | 0/9 (0) |
| HAM | 0/72 (0) | 0/47 (0) | 0/16 (0) | 0/9 (0) |
| Allogeneic SCT, no./no. available (%) |  |  |  |  |
| AlloSCT in CR1 | 0/72 (0) | 0/47 (0) | 0/16 (0) | 0/9 (0) |
| AlloSCT salvage | 0/72 (0) | 0/47 (0) | 0/16 (0) | 0/9 (0) |
| Not done | 72/72 (100) | 47/47 (100) | 16/16 (100) | 9/9 (100) |
| Cause of delay of treatment, no./no. available (%) |  |  |  |  |
| Infection | 22/72 (30.56) | 14/47 (29.79) | 6/16 (37.5) | 2/9 (22.22) |
| Nutritional upbuilding | 0/72 (0) | 0/47 (0) | 0/16 (0) | 0/9 (0) |
| Cost of treatment | 16/72 (22.22) | 8/47 (17.02) | 5/16 (31.25) | 3/9 (33.33) |
| Lack of access | 15/72 (20.83) | 7/47 (14.89) | 5/16 (31.25) | 3/9 (33.33) |
| Unfamiliarity to Treatment | 2/72 (2.78) | 1/47 (2.13) | 0/16 (0) | 1/9 (11.11) |
| Pregnancy | 1/72 (1.39) | 1/47 (2.13) | 0/16 (0) | 0/9 (0) |
| Diagnostic Dilemma | 1/72 (1.39) | 1/47 (2.13) | 0/16 (0) | 0/9 (0) |
| Unknown | 4/72 (5.56) | 4/47 (8.51) | 0/16 (0) | 0/9 (0) |
| No Treatment Delay | 11/72 (15.28) | 11/47 (23.4) | 0/16 (0) | 0/9 (0) |

### Table 2: Treatment outcomes CR/CRi, ED, and 2-year OS of all patients

Note that there are only 2 patients with age > 60 years old. Updated to now include EFS results (some have extremely large values or “Inf” due to insufficient amount of data)

The way EFS is calculated is:

* "Primary refractory disease", "Relapse after first remission", and "Death" are considered events (1).
* "None" is considered event-free (0).

Then taking the proportion of how much of the total was an event.

| AgeGroup | Metric | TDT 0-30 d | TDT 31-60 d | TDT 60+ d |
| --- | --- | --- | --- | --- |
| All Patients | CR\_CRI no./no. available (%) [CI] | 26/47 (55.3%)  [41.2-68.6] | 7/16 (43.8%)  [23.1-66.8] | 4/9 (44.4%)  [18.9-73.3] |
| All Patients | ED  no./no. available (%) [CI] | 19/47 (40.4%)  [27.6-54.7] | 9/16 (56.2%)  [33.2-76.9] | 6/9 (66.7%)  [35.4-87.9] |
| All Patients | OS\_2Y, % [CI] | 2.1%  [0.4-11.1] | 12.5%  [3.5-36] | 0%  [0-29.9] |
| All Patients | EFS, % [CI] | 34/47 (72.3%)  [58.2-83.1] | 11/16 (68.8%)  [44.4-85.8] | 7/9 (77.8%)  [45.3-93.7] |
| Age ≤ 60 y | CR\_CRI no./no. available (%) [CI] | 25/45 (55.6%)  [41.2-69.1] | 6/15 (40%)  [19.8-64.3] | 4/9 (44.4%)  [18.9-73.3] |
| Age ≤ 60 y | ED no./no. available (%) [CI] | 18/45 (40%)  [27-54.5] | 9/15 (60%)  [35.7-80.2] | 6/9 (66.7%)  [35.4-87.9] |
| Age ≤ 60 y  Age ≤ 60 y | OS\_2Y % [CI]  EFS, % [CI] | 2.2%  [0.4-11.6]  32/45 (71.1%) [56.6-82.3] | 6.7%  [1.2-29.8]  11/15 (73.3%) [48-49.1] | 0% [0-29.9]  7/9 (77.8%) [45.3-93.7] |
| Age > 60 y | CR\_CRI no./no. available (%) [CI] | 1/2 (50%)  [9.5-90.5] | 1/1 (100%)  [20.7-100] | NA |
| Age > 60 y | ED no./no. available (%) [CI] | 1/2 (50%)  [9.5-90.5] | 0/1 (0%)  [0-79.3] | NA |
| Age > 60 y  Age > 60 y | OS\_2Y % [CI]  EFS, % [CI] | 0% [0-65.8]  2/2 (100%) [34.2-100] | 100% [20.7-100]  0/1 (0%) [0-79.3] | NA  NA |

### Table 3: ORs for the achievement of CR/CRi, ED, HRs for OS

I am not sure why the OR results of AML and ENL and the CI\_CR of HCTCI are odd. Now includes EFS

| Predictor | OR\_CR | CI\_CR | P\_CR | OR\_ED | CI\_ED | P\_ED | HR\_OS | CI\_OS | P\_OS |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TDT | 1.01 | 0.99-1.03 | 0.439 | 1 | 0-Inf | 1.00 | 0.984 | 0.97-0.999 | 0.036 |
| ENL risk: favorable | 0.00 | 0-Inf | 0.999 | 1 | 0-Inf | 1.00 | 15.811 | 0-Inf | 1.000 |
| ENL risk: intermediate | 5.07e-04 | 0-Inf | 0.999 | 1 | 0-Inf | 1.00 | 1.890 | 0-Inf | 1.000 |
| ENL risk:  unknown | 8.88e-04 | 0-Inf | 0.999 | 1 | 0-Inf | 1.00 | 1.471 | 0-Inf | 1.000 |
| AML types:  AML | 1.07e-03 | 0-Inf | 0.999 | 1 | 0-Inf | 1.00 | 0.981 | 0-Inf | 1.000 |
| AML type: unknown | 8.45e+24 | 0-Inf | 0.996 | 1 | 0-Inf | 1.00 | 0.000 | 0-Inf | 0.999 |
| HCTCI | 4.099e+00 | 0.14-119.85 | 0.413 | 1 | 0-Inf | 1.00 | 0.649 | 0.077-5.449 | 0.690 |
| Age | 9.66e-01 | 0.901-1.034 | 0.326 | 1 | 0-Inf | 1.00 | 1.027 | 0.991-1.065 | 0.147 |
| WBC | 9.99e-01 | 0.991-1.008 | 0.984 | 1 | 0-Inf | 1.00 | 1.00 | 0.995-1.005 | 0.989 |
| LDH | 1.001 | 0.999-1.002 | 0.264 | 1 | 5.676e-58-1.762e+57 | 1.00 | 1.00 | 0.999-1 | 0.271 |
| ECOG | 1.834 | 0.241-13.930 | 0.558 | 1 | 0-Inf | 1.00 | 0.672 | 0.238-1.896 | 0.453 |
| EFS: Death | 0.00 | 0-Inf | 0.999 | 0 | 0-Inf | 0.9999 | 0.940 | 0.109-8.083 | 0.955 |
| EFS: None | 2.65e+15 | 0-Inf | 0.995 | 0 | 0-Inf | 0.9997 | 0.000 | 0-Inf | 0.998 |
| EFS: Primary refractory disease | 0.00 | 0-Inf | 0.997 | 0 | 0-Inf | 0.9997 | 0.176 | 0.043-0.726 | 0.016 |
| EFS: Primary Refractory Disease | 8.86e-01 | 0.066-11.809 | 0.927 | 0 | 0-Inf | 0.9998 | 0.427 | 0.113-1.61 | 0.209 |
| EFS: Relapse after first remission | 2.29e+09 | 0-Inf | 0.997 | 0 | 0-Inf | 0.9997 | 0.000 | 0-Inf | 0.999 |
| EFS: Unknown | 2.78e+09 | 0-Inf | 0.998 | 0 | 0-Inf | 0.9998 | 0.000 | 0-Inf | 0.999 |

## Tests Performed

For the purposes of this analysis, I will not be doing any stratification (breaking up TDT into groups) due to low sample size (difficult to draw conclusions).

So for example, I did not perform the propensity score weighting test since that determines the treatment effects for each treatment TDT group.

### Test 1: Chi-sqaure Test for CR and ED

#### Basic Overview

* **Technical Definition:** The chi-squared test is used to compare observed frequencies of categorical data across groups to expected frequencies under the null hypothesis of no association.
* **What It Is**: This test checks if there’s a relationship between two categories. For example, it can test if remission rates are different for patients who started treatment earlier versus later.
* **How It Was Used:** To compare binary outcomes like complete remission (CR) and early death (ED) between TDT groups.
* **Why It Was Used:** To see if starting treatment sooner or later affected outcomes like remission or early death.
* **Purpose:** To test whether there is a significant association between TDT categories and binary outcomes.
* **Example of Interpreting Results:** If the p-value is small (less than 0.05), it means there’s a strong relationship between the timing of treatment and outcomes. For example, p = 0.169 for remission rates, meaning there was no significant difference between groups.

#### Execution and Analysis

First, we create a contingency table to count how many observations fall into each category

table\_CR <- table(data$TDT, data$TxResponse)

Then we run the chisq-test to determine the p-value:

* The null hypothesis is: The two variables (TDT & Treatment Response) are independent
* The alternative hypothesis is: The two variables are dependent

# Run chi-square test, correct = FALSE for small sample sizes  
chi\_test\_CR <- chisq.test(table\_CR, correct=FALSE)  
  
# Output results  
print(chi\_test\_CR)

##   
## Pearson's Chi-squared test  
##   
## data: table\_CR  
## X-squared = 118.08, df = 135, p-value = 0.8498

#### Results and Conclusions

For complete remission (CR), since p = 0.8498 (greater than 0.05), we fail to reject the null hypothesis. This means there’s no strong evidence that TDT affects Treatment Response

### Test 2: Survival Tests

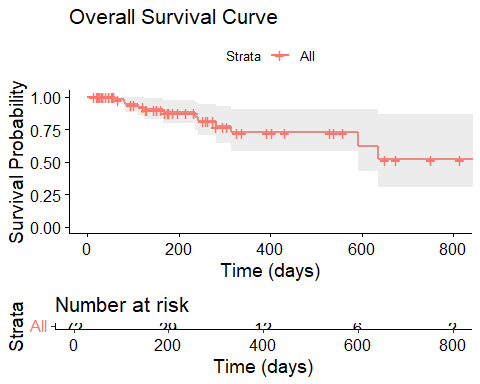
### Test 2a: Kaplan-Meier Survival Curves

This will serve as a setup (visual representation) for the log-ranked test (Test 2b)

#### Basic Overview

* **Technical Definition:** Kaplan-Meier survival curves estimate the probability of survival over time while accounting for censored data (patients lost to follow-up or still alive at the end of the study).
* **What It Is:** This method shows how long patients survive over time. It creates a graph (called a survival curve) that shows the percentage of people still alive at different points in time.
* **Why It Was Used:** To check if patients who started treatment earlier lived longer than those who started later.
* **How It Was Used:** To compare overall survival (OS) across TDT groups. Log-rank tests were used to assess whether survival differences between groups were statistically significant.
* **Purpose:** To visualize and test differences in survival probabilities over time.
* **Example of Interpreting Results:** The survival curves showed no big differences between groups based on TDT. A p-value of 0.211 means there’s no evidence that starting treatment earlier improves survival.

#### Execution and Analysis



#### Results and Conclusions

* Early Phase (0–200 Days)
  + Survival Probability: Starts at 1.00 and declines to approximately 0.80 within the first 200 days.
  + Note that there is a significant decrease in the number at risk between 0 and 200 days compared to the other ones.
  + Interpretation: There is a notable early mortality rate, indicating that a subset of patients experience events (e.g., death or relapse) relatively soon after diagnosis or treatment initiation.
* Mid-Phase (200–600 days)
  + Survival Probability: Continues to decline slowly, but reaches a standstill from 320 to 600 days, reaching around 0.75.
  + Interpretation: There is a gradual reduction in the proportion of patients surviving during this period after a slight decrease/
* Late Decline (600+ days)
  + Survival Probability: The curve flattens out after approximately 625 days and stabilizes around 0.5.
  + Interpretation: The mortality rate decreases substantially beyond this point. Patients who survive past the 625-day mark have a relatively stable long-term survival probability.
* Median Survival Time Estimation: The median survival time (when the survival probability reaches 0.50) appears to be around 625 days.
* Confidence Interval (Gray Shaded Area): A wide confidence interval suggests high variability, particularly at later time points, indicating a limited sample size

### Test 2b: Cox-regression

Note: We did not perform the Log-rank test because I did not split up TDT into groups due to low sample size.

This is also an extension of the survival test performed in test 2a.

#### Basic Overview

* **Technical Definition:** Cox regression models the relationship between survival time and one or more predictors while assuming proportional hazards (i.e., hazard ratios are constant over time).
* **What It Is:** Cox regression looks at how different factors (like age or TDT) affect survival time while accounting for other variables.
* **Why It Was Used:** To see if TDT affects survival when considering other things like patient age or lab results.
* **How It Was Used:** Univariable Cox regression assessed the effect of TDT on OS. Multivariable Cox regression adjusted for confounders like age, WBC, genetic risk, etc.
* **Purpose:** To quantify the effect of predictors (e.g., TDT) on survival outcomes while controlling for other variables.
* **Example of Interpreting Results:** A hazard ratio (HR) tells you how much a factor changes the risk of dying. An HR of 1 means no effect. For ex: an HR for TDT was 1.00 (p=0.617), meaning TDT had no impact on survival.

#### Execution and Analysis

# Fit Cox model (adjust for covariates like Age, WBC, etc.)  
cox\_model <- coxph(  
 Surv(  
 time = SurvivalOS,   
 event = TxResponse == "Primary Refractory Disease")   
 ~ TDT + Age + WBC + ECOG, data = data  
)  
  
# Summarize results  
summary(cox\_model)

## Call:  
## coxph(formula = Surv(time = SurvivalOS, event = TxResponse ==   
## "Primary Refractory Disease") ~ TDT + Age + WBC + ECOG, data = data)  
##   
## n= 72, number of events= 12   
##   
## coef exp(coef) se(coef) z Pr(>|z|)  
## TDT 0.006567 1.006589 0.006765 0.971 0.332  
## Age -0.015707 0.984416 0.025870 -0.607 0.544  
## WBC -0.001031 0.998969 0.005441 -0.190 0.850  
## ECOG -0.969626 0.379225 0.828256 -1.171 0.242  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## TDT 1.0066 0.9935 0.9933 1.020  
## Age 0.9844 1.0158 0.9357 1.036  
## WBC 0.9990 1.0010 0.9884 1.010  
## ECOG 0.3792 2.6370 0.0748 1.923  
##   
## Concordance= 0.592 (se = 0.107 )  
## Likelihood ratio test= 2.15 on 4 df, p=0.7  
## Wald test = 2.01 on 4 df, p=0.7  
## Score (logrank) test = 1.99 on 4 df, p=0.7

# Key outputs:  
# - Hazard Ratios (HR): exp(coef) > 1 indicates increased risk.  
# - Confidence Intervals (CI): HR CI excluding 1 implies significance.  
# - p-values: Variables with p < 0.05 are significant predictors.

#### Results and Conclusions

Key outputs:

* Hazard Ratios (HR): exp(coef) > 1 indicates increased risk.
* Confidence Intervals (CI): HR CI excluding 1 implies significance.
* p-values: Variables with p < 0.05 are significant predictors.
* **No Significant Predictors: All predictors have p > 0.05, meaning none are independently associated with survival.**
* Concordance: 0.592 (moderate ability to rank survival times correctly)
* Global Tests: Likelihood ratio, Wald, and score tests all show p=0.7, confirming the model does not significantly explain survival variation.
* Hazard Ratios:
  + TDT: HR = 1.0066. For every additional hour of TDT, the hazard of death increases by 0.7%, but this is non-significant (p=0.332)
  + Age: HR = 0.9844. For every year increase, the hazard of death decreases by 1.6%, which is non-significant (p=0.544)
  + WBC: HR = 0.9990. For every unit increase in WBC, the hazard of death decreases by 0.1%, which is non-significant (p=0.850)
  + ECOG: HR = 0.3792. This is non-significant (p=0.242)

### Test 3: Logistic Regression

To analyze whether TDT affects early death (ED) or complete remission (CR), while accounting for other covariates like age and WBC, you would use multivariable logistic regression.

#### Basic Overview

* **Technical Definition:** Logistic regression models binary outcomes (e.g., CR or ED) as a function of one or more predictors.
* **What It Is:** Logistic regression predicts outcomes that have two possibilities—like whether a patient achieved remission (yes/no) or died early (yes/no).
* **Why It Was Used:** To check if TDT affects the chances of remission or early death while considering other factors like age and lab results.
* **How It Was Used:** Univariable logistic regression tested associations between TDT and binary outcomes. Multivariable logistic regression adjusted for confounders like age, genetic risk, etc.
* **Purpose:** To estimate odds ratios (ORs) for binary outcomes based on predictors while accounting for other variables.
* **Example on Interpreting Results:** An odds ratio (OR) tells you how much a factor changes the odds of an outcome. measure the strength of association between a categorical outcome (e.g., early death or complete remission) and one or more predictors (e.g., TDT, age, WBC).
  + OR = 1: No effect; the predictor does not change the odds of the outcome. (Ex: an OR of 0.99 (p=0.254) for remission shows that each extra day of TDT didn’t change the odds of remission significantly.)
  + OR > 1: Increased odds; the predictor makes the outcome more likely. (Ex: An OR of 1.5 for TDT means that for every additional day of TDT, the odds of early death increase by 50%.)
  + OR < 1: Decreased odds; the predictor makes the outcome less likely. (Ex: An OR of 0.8 for TDT means that for every additional day of TDT, the odds of early death decrease by 20%.)
  + If OR 1 and p 0.05, TDT significantly affects ED or CR.

#### Execution and Analysis

##### Univariable Analysis

We want to try examining the relationship between TDT & Treatment Response. We do this by first excluding all other variables from the models.

Univariate Logistic Regression model for CR:

# Fit univariate logistic regression model for CR  
cr\_uni\_model <- glm(TxResponse == 'CR' ~ TDT, family = binomial(link = "logit"), data = data)  
  
# 1 = CR.  
  
# Summary of the model  
summary(cr\_uni\_model)

##   
## Call:  
## glm(formula = TxResponse == "CR" ~ TDT, family = binomial(link = "logit"),   
## data = data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.154931 0.327054 0.474 0.636  
## TDT -0.002964 0.006771 -0.438 0.662  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 99.758 on 71 degrees of freedom  
## Residual deviance: 99.564 on 70 degrees of freedom  
## AIC: 103.56  
##   
## Number of Fisher Scoring iterations: 3

# Calculate odds ratio and confidence intervals  
exp(cbind(OddsRatio = coef(cr\_uni\_model), confint(cr\_uni\_model)))

## OddsRatio 2.5 % 97.5 %  
## (Intercept) 1.1675770 0.6154167 2.239575  
## TDT 0.9970409 0.9829259 1.010482

Interpretation:

* TDT Effect: Coefficient = -0.003 (p = 0.662)
* Odds Ratio: 0.997 (95% CI: 0.982 - 1.91) (no effect)
* Interpretation: For each additional day of TDT, the odds of achieving CR decrease by approximately 0.3% (1 - 0.997).
* **Statistical Significance: The effect is not statistically significant (p > 0.05), and the the confidence interval covers 1.0, indicating no effect that TDT has an effect on Treatment response.**

Model Fit:

* AIC: 103.56. Doesn’t suggest a strong model fit in terms explanatory power.
* Residual Deviance: 99.564 on 70 degrees of freedom
  + The null deviance (99.758) and residual deviance (99.564) are nearly identical.
  + A large drop in deviance would indicate that the model improves prediction—but here, the difference is minimal.
  + This suggests that TDT doesn’t explain much of the variability in treatment response (CR).

**Conclusion: TDT does not significant predict treament response (CR)**

##### Multivariable Analysis

Now we consider whether other factors such as age, WBC, etc. affect our conclusions.

First we create the multivariable logistic regression for complete remission (CR)

* We exclude the ED variable from the CR model because it is an outcome variable, not a predictor (vice versa applies later for the ED model)

# Fit logistic regression model for CR  
cr\_multi\_model <- glm(TxResponse == 'CR' ~ TDT + Age + WBC + ECOG,  
 family = binomial(link = "logit"), data = data)  
  
# Summary of the model  
summary(cr\_multi\_model)

##   
## Call:  
## glm(formula = TxResponse == "CR" ~ TDT + Age + WBC + ECOG, family = binomial(link = "logit"),   
## data = data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.3869678 0.8725832 0.443 0.657  
## TDT -0.0041151 0.0069038 -0.596 0.551  
## Age -0.0093407 0.0200046 -0.467 0.641  
## WBC -0.0003477 0.0030943 -0.112 0.911  
## ECOG 0.6738030 0.5617210 1.200 0.230  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 99.758 on 71 degrees of freedom  
## Residual deviance: 97.728 on 67 degrees of freedom  
## AIC: 107.73  
##   
## Number of Fisher Scoring iterations: 4

# Calculate odds ratios and confidence intervals  
exp(cbind(OddsRatio = coef(cr\_multi\_model), confint(cr\_multi\_model)))

## OddsRatio 2.5 % 97.5 %  
## (Intercept) 1.4725091 0.2651032 8.383917  
## TDT 0.9958934 0.9816845 1.009706  
## Age 0.9907027 0.9519980 1.030485  
## WBC 0.9996523 0.9934596 1.005862  
## ECOG 1.9616834 0.6654663 6.179323

Interpretation:

* None of the predictors (TDT, Age, WBC, ECOG) have p-values below 0.05, meaning none are significant.
* Even ECOG, which has the highest odds ratio (1.96), has a p-value of 0.230, indicating it does not significantly predict CR.
* Odds ratios: TDT (0.9959), Age (0.9901), WBC (0.997) all have odds very close to 1. Suggests minimal impact on whether a patients achieves CR.
* Confidence intervals: All the 95% predictors contain 1, indicating that none of them are significant predictors of CR.
* Sample Size Issues: The **small sample size (n=72) severely limits the reliability of the multivariate model.**
* AIC of 107.73 indicates model does not have explanatory power. So we decide to not use the multivatiate model due to sample size

#### Results and Conclusions

**From all the univariate tests, we can conclude that there is no significant association between TDT & Treatment Response (with a potential negative effect but not significant)**

### Optional Check: You may skip this section

I will run a few tests to determine whether it is stable and fits our data appropriately:

# Check VIF for CR model  
# Load required package  
library(car)  
  
# Fit the CR model  
cr\_model <- glm(TxResponse == 'CR' ~ TDT + Age + WBC + ECOG,   
 family = binomial(link = "logit"), data = data)  
  
# Check VIF  
vif(cr\_model)

## TDT Age WBC ECOG   
## 1.034425 1.019667 1.036107 1.032364

# If VIF > 10 - severe multicollinearity  
# If VIF < 5 - generally acceptable

So multicollinearity is not the cause of instability of these models since all VIF’s are < 10.

We can conclude the extreme confidence intervals in multivariate model and odds ratios (OR) are most likely due to sample size.

### Test 4: Restricted Cubic Spline (RCS)

#### Basic Overview

* **Technical Definition:** RCS is a flexible modeling technique that allows for nonlinear relationships between a predictor and an outcome by dividing the predictor into segments joined by “knots.”
* **What It Is:** RCS is a fancy way to look at relationships that aren’t straight lines—it helps find patterns that might be curved or complex.
* **Why It Was Used:** To see if there’s a more complicated relationship between TDT and outcomes like survival or remission.
* **How It Was Used:** To model potential nonlinear relationships between TDT and outcomes like OS or CR without categorizing continuous variables.
* **Purpose:** To avoid loss of information from categorization and better capture complex relationships between variables.

#### Execution and Analysis

Model TDT as a nonlinear predictor. For example, to analyze CR (using 4 knots):

# Load packages  
library(rms)  
library(ggplot2)  
  
# Fit logistic regression model with RCS for TDT  
cr\_rcs\_model <- lrm(TxResponse == 'CR' ~ rcs(TDT, 4) + Age + Male + WBC + ECOG,   
 data = data)  
  
# Print model summary  
print(cr\_rcs\_model)

## Logistic Regression Model  
##   
## lrm(formula = TxResponse == "CR" ~ rcs(TDT, 4) + Age + Male +   
## WBC + ECOG, data = data)  
##   
## Model Likelihood Discrimination Rank Discrim.   
## Ratio Test Indexes Indexes   
## Obs 72 LR chi2 3.61 R2 0.065 C 0.632   
## FALSE 35 d.f. 7 R2(7,72) 0.000 Dxy 0.265   
## TRUE 37 Pr(> chi2) 0.8232 R2(7,54) 0.000 gamma 0.265   
## max |deriv| 2e-07 Brier 0.237 tau-a 0.134   
##   
## Coef S.E. Wald Z Pr(>|Z|)  
## Intercept 0.9893 1.3072 0.76 0.4492   
## TDT -0.0289 0.0795 -0.36 0.7161   
## TDT' -0.0272 0.8299 -0.03 0.9738   
## TDT'' 0.1415 1.6334 0.09 0.9310   
## Age -0.0081 0.0204 -0.39 0.6935   
## Male -0.0895 0.5174 -0.17 0.8626   
## WBC -0.0015 0.0034 -0.46 0.6482   
## ECOG 0.5666 0.5935 0.95 0.3398

# Compare linear vs RCS models using likelihood ratio test  
linear\_model <- lrm(TxResponse == 'CR' ~ TDT + Age + Male + WBC + ECOG,   
 data = data)  
lrtest(cr\_rcs\_model, linear\_model)

##   
## Model 1: TxResponse == "CR" ~ rcs(TDT, 4) + Age + Male + WBC + ECOG  
## Model 2: TxResponse == "CR" ~ TDT + Age + Male + WBC + ECOG  
##   
## L.R. Chisq d.f. P   
## 1.5785953 2.0000000 0.4541637

Interpretation:

* No nonlinear relationships:
  + The RCS terms (TDT, TDT’, TDT’’) are not significant (p > 0.1), and the likelihood ratio test confirms the RCS model does not improve fit over a linear model.
  + All the other predictors **have weak or insignificant effects**, except for ECOG (0.5666) with the strongest effect but no significance.
* Likelihood Ratio Test (chisquare = 1.58, p = 0.45). The RCS model does not significantly improve fit over a linear model.
* R squared value of 0.065 is very low, indicating the model captures little variation in the outcome (CR)
* Moderate Discrimination (C = 0.632): The model weakly distinguishes CR vs. non-CR patients. Slightly better than random guessing but far from ideal (closer to 0.5 is worse, closer to 1 is normal).

#### Results and Conclusions

* **No significant predictors: Age, WBC, Male, or ECOG do not independently affect CR.**
* **TDT’s relationship with CR is weak and does not benefit from this spline transformation. Model performance is weak.**

## Other tests not mentioned in the study

### Test 5: Correlation Analysis:

Assess relationships between continuous variables (e.g., TDT, WBC, LDH). Stronger relationships are higher numbers that are further away from 0. 0 indicates no relationship.

cor\_matrix <- cor(data[, c("TDT", "WBC", "ECOG", "SurvivalEFS", "SurvivalOS")], use = "complete.obs")  
print(cor\_matrix)

## TDT WBC ECOG SurvivalEFS SurvivalOS  
## TDT 1.0000000 -0.12492553 0.12407486 -0.0266559 0.10143550  
## WBC -0.1249255 1.00000000 -0.09276117 -0.1488567 -0.17846969  
## ECOG 0.1240749 -0.09276117 1.00000000 0.1208513 0.06822502  
## SurvivalEFS -0.0266559 -0.14885667 0.12085133 1.0000000 0.88474843  
## SurvivalOS 0.1014355 -0.17846969 0.06822502 0.8847484 1.00000000

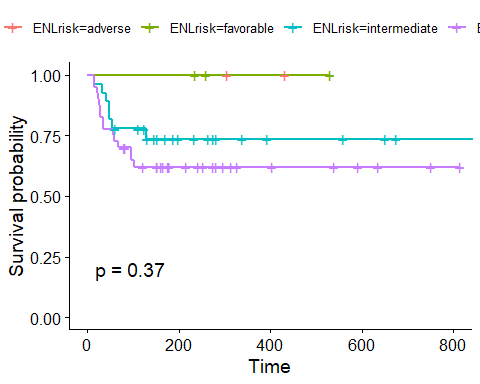
#### Interpretation

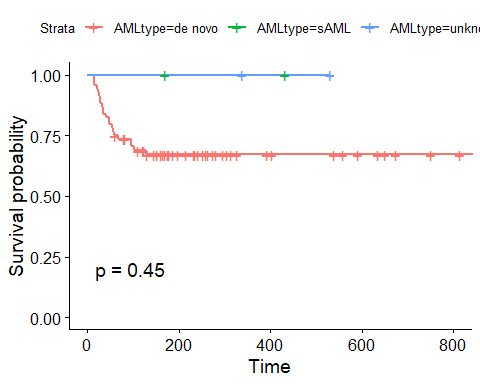
* The strongest relationship is between SurvivalEFS and SurvivalOS. This suggests these measures are tightly linked and may provide complementary information about patient survival.
* **TDT’s almost 0 correlation (-0.03) with SurvivalEFS hints at that there is no relationship between TDT and Survival.**
* Other correlations are relatively weak, which might suggest low interdependence between those variables.

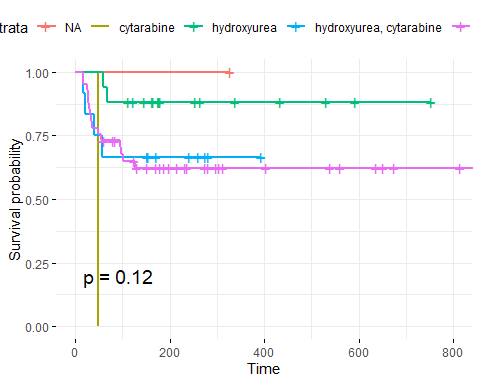
### Test 6: Survival Analysis by Subgroups

Analyze survival curves stratified by categorical variables such as:

* ENL risk group (e.g., favorable, intermediate, adverse, unknown).
* AML type (e.g., de novo vs. secondary AML).
* Use of CytoredAgent (e.g. hydroxyurea, cytarabine, no cytoreduction given)







#### Results and Interpretation

For ENL Risk:

* The green group (ENLrisk = favorable) appears to have the best survival probability over time, followed by the blue group (ENLrisk = intermediate) while the purple group (ENLrisk = unknown) drops faster.
* The p-value (log-rank test) of 0.37 suggests that the differences between these ENL risk levels do not significantly impact survival outcomes.
* There is not enough data to determine survival probability for ENLrisk = adverse
* The lack of statistical significance could be due to small sample size, high variability, or overlapping survival patterns.

For AML Type:

* The red group (AMLtype = de novo), appears to have varying levels of survival probability within the early stages of 0-150 days, but it becomes flattened out from 150+ days onwards with survival probability around 0.65.
* The p-value (log-rank test) of 0.45 suggests that there is no strong evidence that AML type significantly impact survival outcomes.
* There is **not enough data** to determine survival probability for AMLtype = unknown or AMLtype = sAML.

For CytoredAgent:

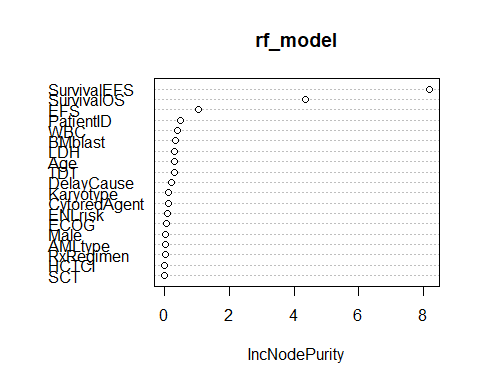
* The green group (hydroxyurea) looks to have the best survival probability overtime followed by the blue group (both hydroxyurea and cytarabine)
* The purple group (CytoredAgent = no cytoreduction given) appears to have the worst survival probability over time
* Log-Rank Test (p-value): The p-value = 0.12 suggests that there is no statistically significant difference in survival between the three groups.

#### Conclusion

**There is no significant difference in survival between the different groups for ENL Risk, AML Type, and Cytored Agent.**

### Test 7: Machine Learning for Feature Importance

Using machine learning (random forests) to determine the most important predictors of survival or treatment response:



**It looks like SurvivalEFS is the most important predictor for survival, followed by: SuvivalOS, EFS, TDT**

### Test 8: LASSO Model

Also try using the LASSO model to determine the key predictors of treatment response:

Definition: LASSO (Least Absolute Shrinkage and Selection Operator) is a special type of regression that helps select the most important variables.

* Standard models (like logistic regression) might include many variables, even if some don’t contribute much.
* LASSO effectively removes unnecessary predictors/factors
* This prevents overfitting (making a model too complex) and improves accuracy

library(glmnet)  
  
# Remove rows with NA values  
data\_clean <- na.omit(data)  
  
# Prepare matrix for LASSO  
x <- model.matrix(TxResponse == 'CR' ~ Age + TDT + WBC + LDH + ECOG + Male, data = data\_clean)  
y <- as.numeric(data\_clean$TxResponse == 'CR')  
  
# Fit LASSO model  
lasso\_model <- cv.glmnet(x, y, family = "binomial", alpha = 1)  
print(lasso\_model)

##   
## Call: cv.glmnet(x = x, y = y, family = "binomial", alpha = 1)   
##   
## Measure: Binomial Deviance   
##   
## Lambda Index Measure SE Nonzero  
## min 0.07994 1 1.419 0.01251 0  
## 1se 0.07994 1 1.419 0.01251 0

#### Results & Interpretation

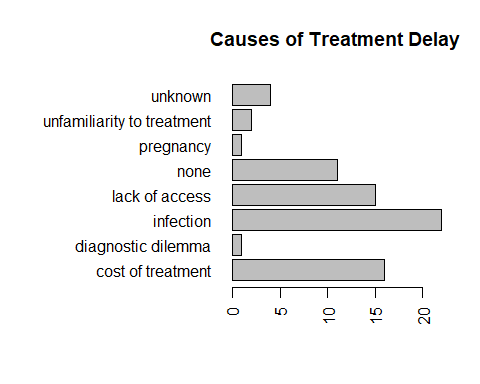
* Measure (Binomial Deviance): Indicates the goodness of fit. Lower values represent better model performance.

Lambda Values (min and 1se):

* Lambda Min (0.07994): Minimizes the binomial deviance, representing the most predictive model, but may include more coefficients.
* Lambda 1-SE (0.07994): A simpler model within one standard error of Lambda Min. This balances accuracy and complexity.
* For both lambda.min and lambda.se, the model selected 0 non-zero coefficients, implying none of the predictors had strong predictive power for treatment response. This might indicate potential issues with variable selection (previous test) or data quality
* Since both selected lambdas resulted in zero predictors, it’s **worth trying alternative models. (unsuccessful)**

### Test 9: Cause of Treatment Delay Analysis

table\_delay <- table(data$DelayCause)  
  
par(mar = c(5, 12, 4, 2))   
barplot(table\_delay, main="Causes of Treatment Delay", horiz = TRUE, cex.names = 1, las = 2)



The **most frequent cause of treatment delay is infection**.

Impact on Outcomes: Assess how different causes of delay affect early death. Using 2 different models:

1. Cox model - for survival analysis: estimates how different predictors (e.g., TDT, ECOG, LDH) affect the likelihood of early death over time. Hazard ratio (HR) tells you if a predictor increases or decreases the risk of an event happening sooner.
2. Logistic Regression - used to predict the probability of an event happening— whether a patient achieves CR (Complete Response) or experiences Early Death (ED). It takes predictors (like age, WBC, ECOG, and delay causes) and estimates their impact on the outcome.

cox\_delay <- coxph(Surv(SurvivalOS, TxResponse == 'Early Death') ~ DelayCause,  
 data = data)  
summary(cox\_delay)

## Call:  
## coxph(formula = Surv(SurvivalOS, TxResponse == "Early Death") ~   
## DelayCause, data = data)  
##   
## n= 72, number of events= 22   
##   
## coef exp(coef) se(coef) z  
## DelayCausediagnostic dilemma -1.878e+01 6.965e-09 1.832e+04 -0.001  
## DelayCauseinfection -4.323e-01 6.490e-01 6.057e-01 -0.714  
## DelayCauselack of access 1.372e-01 1.147e+00 5.775e-01 0.238  
## DelayCausenone 1.792e-01 1.196e+00 6.462e-01 0.277  
## DelayCausepregnancy 1.713e+00 5.545e+00 1.105e+00 1.550  
## DelayCauseunfamiliarity to treatment -1.878e+01 6.965e-09 1.295e+04 -0.001  
## DelayCauseunknown -1.878e+01 6.964e-09 9.234e+03 -0.002  
## Pr(>|z|)  
## DelayCausediagnostic dilemma 0.999  
## DelayCauseinfection 0.475  
## DelayCauselack of access 0.812  
## DelayCausenone 0.782  
## DelayCausepregnancy 0.121  
## DelayCauseunfamiliarity to treatment 0.999  
## DelayCauseunknown 0.998  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## DelayCausediagnostic dilemma 6.965e-09 1.436e+08 0.0000 Inf  
## DelayCauseinfection 6.490e-01 1.541e+00 0.1980 2.127  
## DelayCauselack of access 1.147e+00 8.718e-01 0.3699 3.557  
## DelayCausenone 1.196e+00 8.359e-01 0.3371 4.245  
## DelayCausepregnancy 5.545e+00 1.803e-01 0.6354 48.395  
## DelayCauseunfamiliarity to treatment 6.965e-09 1.436e+08 0.0000 Inf  
## DelayCauseunknown 6.964e-09 1.436e+08 0.0000 Inf  
##   
## Concordance= 0.63 (se = 0.06 )  
## Likelihood ratio test= 8.49 on 7 df, p=0.3  
## Wald test = 3.92 on 7 df, p=0.8  
## Score (logrank) test = 8.3 on 7 df, p=0.3

glm\_delay <- glm(TxResponse == 'Early Death' ~ DelayCause,  
 family = binomial(link="logit"), data=data)  
summary(glm\_delay)

##   
## Call:  
## glm(formula = TxResponse == "Early Death" ~ DelayCause, family = binomial(link = "logit"),   
## data = data)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -0.51083 0.51640 -0.989 0.323  
## DelayCausediagnostic dilemma -17.05524 3956.18036 -0.004 0.997  
## DelayCauseinfection -0.71295 0.72491 -0.984 0.325  
## DelayCauselack of access 0.10536 0.73786 0.143 0.886  
## DelayCausenone -0.04879 0.81211 -0.060 0.952  
## DelayCausepregnancy 18.07689 3956.18036 0.005 0.996  
## DelayCauseunfamiliarity to treatment -17.05524 2797.44199 -0.006 0.995  
## DelayCauseunknown -17.05524 1978.09023 -0.009 0.993  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 88.632 on 71 degrees of freedom  
## Residual deviance: 79.363 on 64 degrees of freedom  
## AIC: 95.363  
##   
## Number of Fisher Scoring iterations: 16

#### Conclusion:

For the first test (coxph):

* All p-values exceed 0.05, meaning no delay factor shows strong statistical significance
* **Pregnancy (p = 0.121) has the largest effect size (HR = 5.545), but it’s not statistically significant**
* Some predictors (Diagnostic dilemma, Unfamiliarity with treatment, Unknown) have standard errors in the thousands, leading to unrealistic HR values near 0 or infinity
* Concordance (C = 0.63, SE = 0.06) suggests moderate discrimination (higher C would indicate stronger model performance).
* Likelihood Ratio Test (p = 0.3), Wald Test (p = 0.8), and Score Test (p = 0.3) all show weak model significance

Conclusion: **DelayCause does not strongly influence early death in this dataset**

For the second test (glm): to determine relationship between survival and cause of delay.

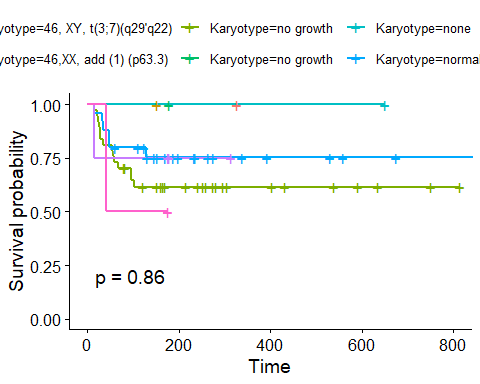
* None of the predictors are statistically significant (p-values all > 0.05), meaning DelayCause does not strongly impact Early Death (CR) in this dataset.
* Residual Deviance (79.36) vs. Null Deviance (88.63) indicates only a minor improvement, meaning DelayCause explains very little of the variation in TxResponse.
* Extreme coefficient values for pregnancy, diagnostic dilemma, unfamiliarity with treatment, and unknown suggest data sparsity in those categories.
* AIC (95.36) suggests the model is not strongly optimized for predicting Early Death.

Conclusion: **DelayCause does not significantly predict Early Death in TxResponse**

### Test 10: Test KaryoType and Response

Compare survival outcomes based on karyotype results (e.g., normal vs. no growth).

km\_fit\_karyo <- survfit(Surv(SurvivalOS, TxResponse == 'Early Death') ~ Karyotype,  
 data = data)  
ggsurvplot(km\_fit\_karyo, data=data, pval=TRUE)



Also perform chi-square test for Karyotype and response

table\_karyo\_response <- table(data$Karyotype, data$TxResponse == 'CR')  
chisq.test(table\_karyo\_response)

##   
## Pearson's Chi-squared test  
##   
## data: table\_karyo\_response  
## X-squared = 4.984, df = 7, p-value = 0.6619

Conclusion

From the plot, it looks like **karyotype = normal has the best survival probability followed by no growth, but it is not significant (p = 0.986).**

From Chisquare test, since p-value is 0.6619, it is **not significant.**

## Summary: What Did All These Tests Show?

To answer one big question: Does starting treatment sooner improve patient outcomes? Here’s what they found:

* Patients who started treatment later didn’t have worse remission rates or survival times compared to those who started earlier.
* Statistical tests confirmed that TDT didn’t significantly affect outcomes like remission, early death, or overall survival—even after adjusting for other factors like age and lab results.
* Study limitations: Small sample size

These findings suggest it might be okay to wait for important test results before starting treatment in some cases—especially if patients are stable—so doctors can choose the best possible treatment plan for each individual.

## Results and Conclusions

The study investigated whether Treatment Delay Time (TDT) influences prognosis in Acute Myeloid Leukemia (AML). After extensive statistical analysis, including logistic regression, Cox proportional hazards models, Kaplan-Meier survival curves, and machine learning techniques such as LASSO and random forests, the findings indicate that TDT does not significantly impact key clinical outcomes. Specifically, there was no strong statistical association between TDT and complete remission (CR), early death (ED), or overall survival (OS). Models assessing TDT alongside other potential prognostic variables, such as age, white blood cell count, lactate dehydrogenase levels, and performance status, demonstrated that these factors had more relevance in predicting AML outcomes than treatment delay.

*Implications for Practice*

The implications of these findings suggest that delaying induction therapy in some patients, particularly those requiring additional diagnostic assessments or preparatory stabilization, may not necessarily compromise treatment success. In clinical practice, this means physicians may have more flexibility in initiating treatment without fearing adverse consequences solely due to delay. However, barriers to timely treatment initiation, such as financial constraints, accessibility issues, and infections, remain a critical concern in patient management. While TDT itself may not be a direct determinant of prognosis, addressing these systemic delays is essential for optimizing care delivery.

*Study Limitations*

Despite these findings, the study has limitations. The relatively small sample size may reduce statistical power, potentially obscuring subtle effects of TDT that might emerge in a larger dataset.

*Recommendations for Future Research*

For future research, expanding the sample size would improve generalizability and statistical robustness. Investigating the role of genetic mutations and cytogenetic abnormalities alongside TDT may offer deeper insights into disease mechanisms. Additionally, stratifying patients based on risk groups or disease subtypes could reveal whether certain populations are more affected by treatment delays than others.   
  
Overall, the study does not support a direct correlation between TDT and AML prognosis.