

Logistic Modeling of Adult Acute Myeloid Leukemia Relapses with Clinical and Disease Characteristics

Abstract

The incidence of adult acute myeloid leukemia (AML) continues to rise and currently, this blood pathology accounts for over 10,000 deaths among US adults each year. In this paper, we sought to model the likelihood that an individual diagnosed with AML will relapse after a complete remission is achieved using logistic modeling methodologies. Initially, a large host of factors from clinical and disease data were investigated, but subsequent analyses later revealed that age at diagnosis and cytogenetic risk level were the most influential factors contributing significant predictive information to the model. The logistic model developed here will likely be of value to the oncology field and those afflicted with AML.

Introduction

Acute myeloid leukemia (AML) is the most common type of acute leukemia among US adults, characterized by the infiltration of cancer cells in the hematopoietic system, and accounting for over 10,000 deaths in the US last year alone [1]. Currently, most individuals diagnosed with AML receive high-dose chemotherapy in an attempt to achieve a complete remission, indicating there is no evidence of disease [2]. However, despite this rigorous treatment regimen, for a large majority of patients, their AML will recur, meaning that undetected, residual disease persists and progresses so that it is once again clinically appreciable [3]. Predicting which AML patients are most likely to experience a relapse or recurrence of their disease could have tremendous prognostic value and aid physicians in treatment decisions. This project sought to model the likelihood that the AML of a diagnosed adult will relapse once complete remission is achieved based on clinical and tumor characteristics. Although there are numerous tumor and clinical characteristics, many of which are out of the scope of this analysis. Therefore, the following factors were investigated: age at diagnosis, sex, peripheral blood and bone marrow myeloblast count, white blood cell count, cytogenetic and molecular risk levels, and AML subtypes of the French-American-British (FAB) classification system. Clinical and tumor data from a 2013 *New England Journal of Medicine* study was utilized and publicly accessible from The Cancer Genome Atlas [4].

Methodology

AML recurrence greatly diminishes the prognosis of a patient and is likely influenced by a host of factors. To investigate how the aforementioned factors influence the likelihood of AML recurrence, we mined publicly-accessible data from The Cancer Genome Atlas for patients enrolled in a 2013 study, $n=200$, but due to the incompleteness of some patient data, our sample size was reduced to 175 patients. After the raw data was downloaded, we performed further data processing in accord with the parameters required for logistic modeling. We formatted the response variable, disease response (y), as a binary variable, coding it as a dummy variable with $y=1$ if the patient was disease free and $y=0$ if the patient's disease recurred. Accordingly, we transformed the non-continuous variables (FAB, cytogenetic risk level and molecular risk level) to be interpreted as categorical variables.

We then generated a generalized logistic model, using all variables mentioned above as factors influencing disease response. To determine the overall adequacy of this model, a Chi-square test was performed, with the following null hypothesis, $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ where $k=17$, and the following alternative hypothesis, H_a : at least one β_i is different from zero for $i=1, 2, \dots, k$. After computing the Chi-square test for the complete logistic model, we determined that additional models would likely more adequately model the data of interest. As a result, model selection was performed using stepwise elimination in the backward direction to determine which, of all the embedded models, best approximates the given data. Optimal selection of the returned models was based upon the Akaike Information Criterion (AIC) results under the assumption that the model having the lowest AIC is the best model. As some models were considered indistinguishable (AIC difference < 2), we opted to select the most parsimonious of models, thereby eliminating factors that didn't contribute to this model. A subsequent Chi-square test was then computed for the newly selected model. Further, summary data returned from the refined model indicate that both remaining factors contribute significant information to the response variable. Wald confidence intervals for the parameter coefficients of the refined model were then computed. We also sought to visualize our refined model and did so by plotting the response variable by each respective predictor value. Last, we utilized the refined model, which predicts disease response according to age and cytogenetic risk level, to predict the odds of an individual staying disease free at different values of these factors.

Analysis

Initially, a complete logistic model, including all factors, was employed to predict the binomial response variable, AML disease response. Yet, upon closer examination of this model with the Chi-square test and the t-tests of coefficient estimates, we inferred that additional models may better model the given data, although the complete model was technically adequate for modeling the data with only one coefficient estimate less than 0.05, thereby allowing us to reject the null hypothesis.

Nevertheless, we performed stepwise model selection using backward elimination to arrive at the best model. As the simplest of models and one with an AIC of 234.23, we selected the logistic model with patient age at diagnosis and cytogenetic risk levels as the contributing factors for predicting AML disease response. Further inspection of the adequacy of the refined model with a Chi-square test revealed that at least one coefficient estimate was different from zero ($p < 0.005, 0.05$), thereby allowing us to reject the null hypothesis and conclude that the refined model is adequate for predicting disease response. Moreover, additional t-tests of the different factors, as shown in the summary data of the refined model, suggest that both predictor variables, age and cytogenetic risk, are significant in explaining disease response with verification from the resulting p-values of 0.04224 and 0.00527, respectively.

The refined logistic model equation can be expressed as: $\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Adding the coefficient estimates produces the following model equation: $\ln \frac{\pi}{1-\pi} = -0.545 - 1.073x_1 + 0.134x_2 + 0.021x_3$. Considering the nature of the data, there is no reason to interpret the coefficient of the intercept as no individual of age zero will be diagnosed with adult AML. Alternatively, interpreting the other coefficient estimates may provide valuable insight into AML disease response. For each additional year of age at diagnosis (x_1), the log odds of staying disease free (not recurring) increases by 0.021, holding cytogenetic risk level constant. Being placed in the intermediate cytogenetic risk level, compared to the poor cytogenetic risk level, decreases the log odds of staying disease free by 1.073, holding age constant. Conversely, being placed in the good cytogenetic risk level, compared to the poor cytogenetic risk level, increases the log odds of staying disease free by 0.134, holding age constant. Confidence intervals show with 95% confidence that the true value of the slope coefficient for age is between 0.00074 and 0.0416, for the intermediate cytogenetic risk level is between -1.826 and -0.3192, and for good cytogenetic risk level is between -0.8324 and 1.1001, which all coincide with the coefficient estimates in the model equation.

Results

In general, the age at diagnosis and cytogenetic risk level contributed the most significant predictive information for AML disease response. A large host of factors were initially considered in our model, but after numerous tests and model selection were performed, many variables were excluded due to their lack of significance. Accordingly, the best fitting model, which included the two factors (three dummy variables) was selected due to the AIC score, Chi-square and t- test results.

The logistic model generated and selected here affords an individual with the ability to predict the likelihood of an AML patient staying disease free versus relapsing by considering two factors. We sought to put our logistic model to the test for predicting the chance of staying disease free of two hypothetical AML patients, whom were diagnosed at the ages of 35 and 70, at each respective cytogenetic risk level. At the age of 35, our fitted model predicted that the probability an individual stays disease free in the poor risk level to be 54.9%, in the intermediate level to be 29.4%, and in the good level to be 58.2%. Comparably, the fitted model predicts the 70-year-old to fare much better with a 71.8% likelihood in the poor level, 46.6% likelihood in the intermediate level, and 74.5% likelihood in the good level of staying disease free. The influence of age at diagnosis may seem contrary to intuition considering that younger patients typically can endure longer courses of high-dose chemotherapy due to a more rapidly regenerating hematopoietic system, thereby increasing their overall survival rate [2]. However, it is notable to clarify that this model predicts for the likelihood of staying AML free, *only* once a complete remission of disease has already been achieved. Therefore, older patients may, in fact, have a better likelihood of staying disease free if a complete remission is achieved, thus potential criticisms of the effect of age as an artifact can be dismissed.

The fitted logistic model generated for this project was based upon publicly-accessible data of adult AML patients stored in The Cancer Genome Atlas. As this model is based upon the patient data from a single study, future work could incorporate additional data from past and prospective studies alike to increase the robustness of this model. Furthermore, while this model may be of less relevance and value to the general population, the unfortunate truth is that with the American Cancer Society predicting over 20,000 new cases of adult AML to occur in 2017 alone, many patients, family members, friends and professionals in the field of oncology, would likely find significant value in the utility of this model [5].

References

1. De Kouchkovsky I, Abdul-hay M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood Cancer J*. 2016;6(7):e441.
2. Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl J Med*. 2015;373(12):1136-52.
3. Szer J. The prevalent predicament of relapsed acute myeloid leukemia. *Hematology Am Soc Hematol Educ Program*. 2012;2012:43-8.
4. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059-74.
5. Adult Acute Myeloid Leukemia Treatment. www.cancer.gov/types/leukemia/hp/adult-aml-treatment. Web. 14 April 2017.

Appendix

RStudio Code

```
# Import dataset
> data <- read.csv(file.choose())
# Convert FAB, Cyogenetic and molecular risk level to categorical variables
data$FAB <- factor(data$FAB)
data$Cyto <- factor(data$Cyto)
data$Molec <- factor(data$Molec)
# Fit logistic model including all parameters
> logfitx <- glm(DFS ~ AGE+FAB+BM.Blast+PB.Blast+Cyto+Molec+Sex+WBC, data=data,
family="binomial")
> summary(logfitx)
# Test the overall adequacy of the model with Chi-Square test
> anova(logfitx, test="Chisq")
# Perform model selection with backward selection
> summary(step(logfitx, direction="backward"))
# Fit new logistic model selected by backward selection
> logfitx4 <- glm(data$DFS ~ data$Cyto + data$AGE, family="binomial")
> summary(logfitx4)
# Test the overall adequacy of the selected model
> anova(logfitx4, test="Chisq")
# Investigate parameter coefficients of selected model
> exp(coef(logfitx4))
# Compute Wald confidence intervals using standard error
> confint.default(logfitx4)
# Display variable average
> mean(data$AGE)
# Generate log plots
> plot(jitter(data$DFS, factor=0.5)~data$Cyto, ylab="Disease Response", xlab="Cytogenetic Risk Level")
> plot(jitter(data$DFS, factor=0.5)~data$AGE, ylab="Disease Response", xlab="Age at Diagnosis")
# Make predictions for AML-diagnosed 35 year-old with fitted model by cytogenetic risk level
# For poor cytogenetic risk level
> pp_poor35 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0+coef(logfitx4)[3]*0 +coef(logfitx4)[4]*35)
> pred_poor35<-exp(pp_poor35)/(1+exp(pp_poor35))
> pred_poor35
# For intermediate cytogenetic risk level
> pp_int35 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*1+coef(logfitx4)[3]*0 +coef(logfitx4)[4]*35)
```

```

> pred_int35<-exp(pp_int35)/(1+exp(pp_int35))
> pred_int35
# For good cytogenic risk level
> pp_good35 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0+coef(logfitx4)[3]*1 +coef(logfitx4)[4]*35)
> pred_good35<-exp(pp_good35)/(1+exp(pp_good35))
> pred_good35
# Make predictions for AML-diagnosed 70 year-old with fitted model by cytogenic risk level
# For poor cytogenic risk level
> pp_poor70 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0+coef(logfitx4)[3]*0 +coef(logfitx4)[4]*70)
> pred_poor70 <-exp(pp_poor70)/(1+exp(pp_poor70))
> pred_poor70
# For intermediate cytogenic risk level
> pp_int70 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*1+coef(logfitx4)[3]*0 +coef(logfitx4)[4]*70)
> pred_int70<-exp(pp_int70)/(1+exp(pp_int70))
> pred_int70
# For good cytogenic risk level
> pp_good70 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0+coef(logfitx4)[3]*1 +coef(logfitx4)[4]*70)
> pred_good70<-exp(pp_good70)/(1+exp(pp_good70))
> pred_good70

```

RStudio Output

```

> data <- read.csv(file.choose())
> head(data)
  AGE DFS FAB BM.Blast PB.Blast Cyto Molec Sex WBC
1  52  1 M2    67    18  2  2 Female 4.1
2  61  1 M3    73    68  2  2 Female 86.4
3  55  1 M4    35     8  2  2 Male 22.2
4  74  1 M3    51     0  2  2 Male 31.5
5  42  1 M3    88     2  2  2 Male  2.1
6  43  1 M2    40    39  2  2 Male  4.3

> data$FAB <- factor(data$FAB)
> data$Cyto <- factor(data$Cyto)
> data$Molec <- factor(data$Molec)

> summary(logfitx)
Call:
glm(formula = dat$DFS ~ dat$FAB + dat$Cyto + dat$Molec + dat$PB.Blast +
  dat$BM.Blast + dat$WBC + dat$Sex + dat$AGE, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9453 -1.0262 -0.3967  1.0327  1.8213

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.277e+00  1.268e+00  1.007  0.3138

```

```

dat$FABM1 -6.721e-01 6.633e-01 -1.013 0.3109
dat$FABM2 -5.811e-01 6.937e-01 -0.838 0.4022
dat$FABM3 -1.085e+00 9.858e-01 -1.100 0.2711
dat$FABM4 6.450e-02 6.914e-01 0.093 0.9257
dat$FABM5 -1.087e+00 8.077e-01 -1.346 0.1784
dat$FABM6 1.697e+01 1.355e+03 0.013 0.9900
dat$FABM7 9.292e-01 1.574e+00 0.590 0.5550
dat$FABnc -1.501e+01 2.400e+03 -0.006 0.9950
dat$Cyto1 1.313e+00 8.225e-01 1.597 0.1104
dat$Cyto2 2.911e-01 1.748e+00 0.167 0.8677
dat$Molec1 1.494e-01 7.567e-01 0.197 0.8435
dat$Molec2 -2.424e-02 1.801e+00 -0.013 0.9893
dat$PB.Blast 8.970e-03 7.875e-03 1.139 0.2547
dat$BM.Blast -5.741e-03 1.080e-02 -0.531 0.5952
dat$WBC -2.614e-03 3.968e-03 -0.659 0.5100
dat$SexMale -2.811e-01 3.483e-01 -0.807 0.4197
dat$AGE -2.571e-02 1.186e-02 -2.168 0.0302 *

```

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 242.55 on 174 degrees of freedom
Residual deviance: 209.73 on 157 degrees of freedom
AIC: 245.73

Number of Fisher Scoring iterations: 15

> anova(logfitx, test="Chisq")

Analysis of Deviance Table

Model: binomial, link: logit

Response: DFS

Terms added sequentially (first to last)

| | Df | Deviance | Resid. Df | Resid. Dev | Pr(>Chi) |
|----------|----|----------|-----------|------------|-------------|
| NULL | | | 174 | 242.55 | |
| AGE | 1 | 3.3666 | 173 | 239.18 | 0.066531 . |
| FAB | 8 | 15.2105 | 165 | 223.97 | 0.055179 . |
| BM.Blast | 1 | 0.0479 | 164 | 223.93 | 0.826748 |
| PB.Blast | 1 | 0.2544 | 163 | 223.67 | 0.613971 |
| Cyto | 2 | 12.8421 | 161 | 210.83 | 0.001627 ** |
| Molec | 2 | 0.0283 | 159 | 210.80 | 0.985934 |
| Sex | 1 | 0.6404 | 158 | 210.16 | 0.423570 |
| WBC | 1 | 0.4308 | 157 | 209.73 | 0.511588 |

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

> summary(step(logfitx, direction="backward"))

Start: AIC=245.73

DFS ~ AGE + FAB + BM.Blast + PB.Blast + Cyto + Molec + Sex +
WBC

| | Df | Deviance | AIC |
|------------|----|----------|--------|
| - Molec | 2 | 209.78 | 241.78 |
| - FAB | 8 | 222.65 | 242.65 |
| - BM.Blast | 1 | 210.01 | 244.01 |
| - WBC | 1 | 210.16 | 244.16 |
| - Sex | 1 | 210.38 | 244.38 |
| - Cyto | 2 | 212.71 | 244.71 |
| - PB.Blast | 1 | 211.04 | 245.04 |
| <none> | | 209.73 | 245.73 |
| - AGE | 1 | 214.62 | 248.62 |

Step: AIC=241.78

DFS ~ AGE + FAB + BM.Blast + PB.Blast + Cyto + Sex + WBC

| | Df | Deviance | AIC |
|------------|----|----------|--------|
| - FAB | 8 | 222.80 | 238.80 |
| - BM.Blast | 1 | 210.10 | 240.10 |
| - WBC | 1 | 210.19 | 240.19 |
| - Sex | 1 | 210.43 | 240.43 |
| - PB.Blast | 1 | 211.12 | 241.12 |
| <none> | | 209.78 | 241.78 |
| - AGE | 1 | 214.79 | 244.79 |
| - Cyto | 2 | 222.91 | 250.91 |

Step: AIC=238.8

DFS ~ AGE + BM.Blast + PB.Blast + Cyto + Sex + WBC

| | Df | Deviance | AIC |
|------------|----|----------|--------|
| - Sex | 1 | 222.85 | 236.85 |
| - WBC | 1 | 223.20 | 237.20 |
| - BM.Blast | 1 | 224.33 | 238.33 |
| <none> | | 222.80 | 238.80 |
| - PB.Blast | 1 | 224.98 | 238.98 |
| - AGE | 1 | 225.98 | 239.98 |
| - Cyto | 2 | 236.95 | 248.95 |

Step: AIC=236.85

DFS ~ AGE + BM.Blast + PB.Blast + Cyto + WBC

| | Df | Deviance | AIC |
|------------|----|----------|--------|
| - WBC | 1 | 223.24 | 235.24 |
| - BM.Blast | 1 | 224.45 | 236.45 |
| <none> | | 222.85 | 236.85 |

- PB.Blast 1 225.12 237.12
- AGE 1 226.00 238.00
- Cyto 2 237.05 247.05

Step: AIC=235.24

DFS ~ AGE + BM.Blast + PB.Blast + Cyto

| | Df | Deviance | AIC |
|------------|----|----------|--------|
| - BM.Blast | 1 | 224.99 | 234.99 |
| - PB.Blast | 1 | 225.12 | 235.12 |
| <none> | | 223.24 | 235.24 |
| - AGE | 1 | 226.47 | 236.47 |
| - Cyto | 2 | 237.07 | 245.07 |

Step: AIC=234.99

DFS ~ AGE + PB.Blast + Cyto

| | Df | Deviance | AIC |
|------------|----|----------|--------|
| - PB.Blast | 1 | 226.23 | 234.23 |
| <none> | | 224.99 | 234.99 |
| - AGE | 1 | 227.96 | 235.96 |
| - Cyto | 2 | 237.77 | 243.77 |

Step: AIC=234.23

DFS ~ AGE + Cyto

| | Df | Deviance | AIC |
|--------|----|----------|--------|
| <none> | | 226.23 | 234.23 |
| - AGE | 1 | 230.50 | 236.50 |
| - Cyto | 2 | 239.18 | 243.18 |

Call: glm(formula = DFS ~ AGE + Cyto, family = "binomial", data = data)

Coefficients:

| (Intercept) | AGE | Cyto1 | Cyto2 |
|-------------|---------|----------|---------|
| -0.54529 | 0.02119 | -1.07282 | 0.13389 |

Degrees of Freedom: 174 Total (i.e. Null); 171 Residual

Null Deviance: 242.6

Residual Deviance: 226.2 AIC: 234.2

```
> logfitx4 <- glm(data$DFS ~ data$Cyto + data$AGE, family="binomial")  
> summary(logfitx4)
```

Call:

glm(formula = data\$DFS ~ data\$Cyto + data\$AGE, family = "binomial")

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -1.7008 | -1.0684 | 0.7396 | 1.0450 | 1.6986 |

Coefficients:

```
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.54529   0.66912  -0.815  0.41511
data$Cyto1  -1.07282   0.38451  -2.790  0.00527 **
data$Cyto2   0.13389   0.49300   0.272  0.78594
data$AGE     0.02119   0.01043   2.031  0.04224 *
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 242.55 on 174 degrees of freedom
Residual deviance: 226.23 on 171 degrees of freedom
AIC: 234.23

Number of Fisher Scoring iterations: 4

```
> anova(logfitx4, test="Chisq")
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: data\$DFS

Terms added sequentially (first to last)

```
      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                174    242.55
data$Cyto  2  12.0492    172    230.50 0.002419 **
data$AGE   1   4.2719    171    226.23 0.038747 *
```

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

```
> exp(coef(logfitx4))
```

```
(Intercept) data$Cyto1 data$Cyto2 data$AGE
 0.5796759  0.3420436  1.1432675  1.0214166
```

```
> exp(coef(logfitx4))-1
```

```
((Intercept) data$Cyto1 data$Cyto2 data$AGE
-0.42032413 -0.65795635 0.14326751 0.02141662
```

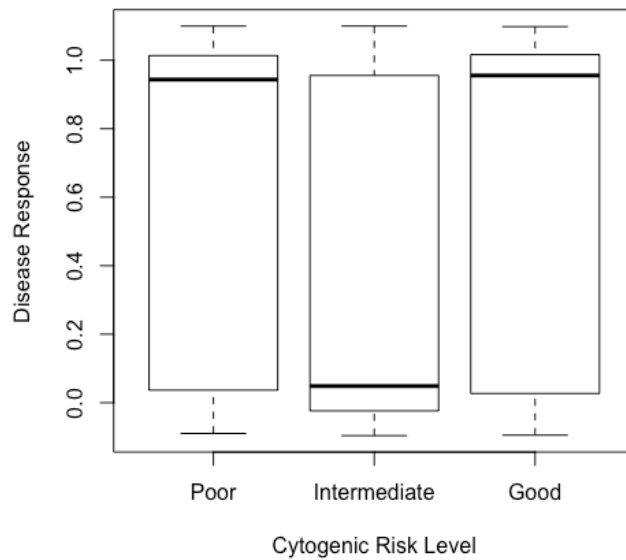
```
> confint.default(logfitx4)
```

```
  2.5 %   97.5 %
(Intercept) -1.8567343734 0.76616203
data$Cyto1  -1.8264377497 -0.31919609
data$Cyto2  -0.8323759238  1.10015673
data$AGE    0.0007430819  0.04163793
```

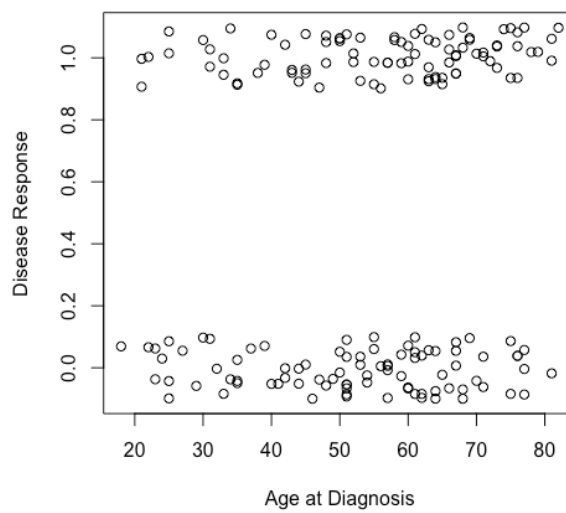
```
> mean(data$AGE)
```

```
[1] 54.33143
```

```
> data$Cyto = factor( data$Cyto, levels = 0:2, abels = c("Poor", "Intermediate", "Good"))
> plot(jitter(data$DFS, factor=0.5)~data$Cyto, ylab="Disease Response", xlab="Cytogenic Risk Level")
```



```
> plot(jitter(data$DFS, factor=0.5)~data$AGE, ylab="Disease Response", xlab="Age at Diagnosis")
```



```
# Make predictions for AML-diagnosed 35 year-old with fitted model by cytogenetic risk level
```

```
# For poor cytogenetic risk level
```

```
> pp_poor35 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0  
+           +coef(logfitx4)[3]*0 +coef(logfitx4)[4]*35)
```

```
> pred_poor35<-exp(pp_poor35)/(1+exp(pp_poor35))
```

```
> pred_poor35
```

```
(Intercept)
```

```
0.5489382
```

```
# For intermediate cytogenetic risk level
```

```
> pp_int35 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*1  
+           +coef(logfitx4)[3]*0 +coef(logfitx4)[4]*35)
```

```
> pred_int35<-exp(pp_int35)/(1+exp(pp_int35))
```

```
> pred_int35
```

```
(Intercept)
```

```
0.293917
```

```
# For good cytogenetic risk level
```

```
> pp_good35 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0  
+           +coef(logfitx4)[3]*1 +coef(logfitx4)[4]*35)
```

```
> pred_good35<-exp(pp_good35)/(1+exp(pp_good35))
```

```
> pred_good35
```

```
(Intercept)
```

```
0.5818255
```

```
# Make predictions for AML-diagnosed 70 year-old with fitted model by cytogenetic risk level
```

```
# For poor cytogenetic risk level
```

```
> pp_poor70 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0  
+           +coef(logfitx4)[3]*0 +coef(logfitx4)[4]*70)
```

```
> pred_poor70 <-exp(pp_poor70)/(1+exp(pp_poor70))
```

```
> pred_poor70
```

```
(Intercept)
```

```
0.7187054
```

```
# For intermediate cytogenetic risk level
```

```
> pp_int70 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*1  
+           +coef(logfitx4)[3]*0 +coef(logfitx4)[4]*70)
```

```
> pred_int70<-exp(pp_int70)/(1+exp(pp_int70))
```

```
> pred_int70
```

```
(Intercept)
```

```
0.466359
```

```
# For good cytogenetic risk level
```

```
> pp_good70 <- (coef(logfitx4)[1] + coef(logfitx4)[2]*0  
+           +coef(logfitx4)[3]*1 +coef(logfitx4)[4]*70)
```

```
> pred_good70<-exp(pp_good70)/(1+exp(pp_good70))
```

```
> pred_good70
```

```
(Intercept)
```

```
0.7449656
```