```
setwd("/Users/Chang/Desktop")

mydata1 <- read.csv("imout1.csv")

mydata1$group=as.factor(mydata1$group)

mydata1$HF=as.factor(mydata1$HF)

mydata1$COPD=as.factor(mydata1$COPD)

mydata1$CAD=as.factor(mydata1$CAD)

mydata1$O2new=as.factor(mydata1$O2new)

summary(mydata1)

View(mydata1)

my_imp1<- mice(mydata1, m=5, method =c("","","","pmm","pmm","",""), seed=500)

summary(mydata1$Feritin)

summary(mydata1$DDimer)

my_impimpFeritin

final_data1<-complete(my_imp1,1)

write.csv(final_data1,file = "/Users/Chang/Desktop/impute1.csv")

stripplot(my_imp1, pch = 20, cex = 1.2)
```
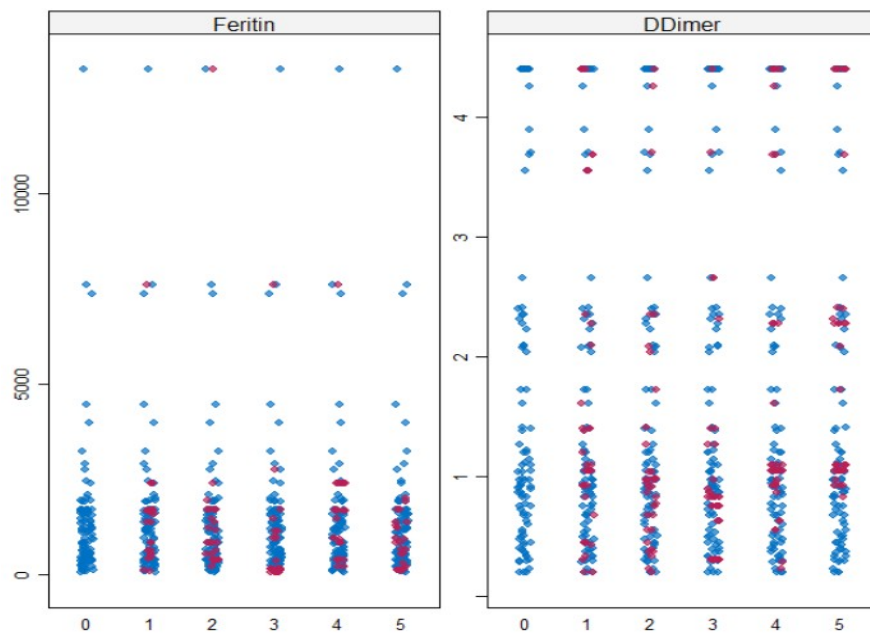
**Option1:**

**HF CAD COPD Ferritin DDimer O2**

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | -4.1961 | 0.7404 | 32.1172 | <.0001 |
| HF | 1 1 | -3.1565 | 1.8452 | 2.9261 | 0.0872 |
| CAD | 1 1 | 0.3249 | 1.2535 | 0.0672 | 0.7955 |
| COPD | 1 1 | -0.9089 | 1.1409 | 0.6346 | 0.4257 |
| Feritin | 1 | 0.000627 | 0.000224 | 7.8298 | 0.0051 |
| DDimer | 1 | 0.4152 | 0.2051 | 4.0967 | 0.0430 |
| O2new | 1 1 | 2.8472 | 0.6109 | 21.7188 | <.0001 |

**Odds Ratio Estimates**

| Effect | Point Estimate | 95% Wald Confidence Limits | |
|---|---|---|---|
| HF 1 vs 0 | 0.043 | 0.001 | 1.584 |
| CAD 1 vs 0 | 1.384 | 0.119 | 16.148 |
| COPD 1 vs 0 | 0.403 | 0.043 | 3.771 |
| Feritin | 1.001 | 1.000 | 1.001 |
| DDimer | 1.515 | 1.013 | 2.264 |
| O2new 1 vs 0 | 17.239 | 5.206 | 57.088 |

**Association of Predicted Probabilities and Observed Responses**

| | | | |
|---|---|---|---|
| Percent Concordant | 87.8 | Somers' D | 0.757 |
| Percent Discordant | 12.2 | Gamma | 0.757 |
| Percent Tied | 0.0 | Tau-a | 0.272 |
| Pairs | 2970 | c | 0.878 |

**Hosmer and Lemeshow Goodness-of-Fit Test**

| Chi-Square | DF | Pr > ChiSq |
|---|---|---|
| 12.7747 | 8 | 0.1198 |

e. Logistic regression analysis for the risk of intubation: What were the variables included into calculation (name all variables). How did the authors prevent overfitting of the model? Provide R2 and the p-value! (see also g.)

HF CAD COPD Ferritin DDimer O2

Hosmer and Lemeshow Goodness of Fit test:

P=0.1198>0.05

Therefore, the model is good fit the data (prevent overfitting of the model).

C=0.878 This is AUC It also shows the model is good.

The odds ratio of Invasive Mechanical Ventilation is 17.239, comparing O2 sat<90% to O2 sat>=90%, adjusted for other covariates. (95% Confidence interval: (5.206,57.088))

The odds ratio of Invasive Mechanical Ventilation is

1.001, when Ferritin increase 1 unit, adjusted for other covariates. (95% Confidence interval: (1.000, 1.001))

The odds ratio of Invasive Mechanical Ventilation is 1.515, when Ddimer increase 1 unit, adjusted for other covariates. (95% Confidence interval: (5.206, 57.088))

g. As usual in retrospective studies, there are a lot of missing values. Please clarify how you handled missing data especially for the regression analysis and provide the number of valid cases for the regression calculation.

Using multiple imputation to handle missing data.

Calculates imputations for univariate missing data by predictive mean matching (Ferritin and Ddimer).