STAT434_FinalProject

Kexin Yang(kexiny5)

2020/12/14

Introduction

COVID-19 pandemic dates back to its emergence in December 2019 in Wuhan China and lately ravages the global. COVID-19 is a highly infectious disease, which might cause severe acute respiratory illness and even mortality. According to World Health Organization(WHO), there have been more than 70 million comfirmed cases of COVID-19 including more than 1.5 million deaths globally. It is important to identify the risk factors for the disease severtiy and death as more and more people are getting infected and the health care system is getting overwhelmed. It is necessary to know which groups of the patients are more susceptible to COVID-19. In June 25, 2020, CDC updated the expanded list of people at risk of severe COVID-19 illness. It warns that the the probability of an adult having severe illness increases steadily as aging. Also, the underlying medical conditions including chronic kidney disease, type 2 diabetes and serious heart conditions may escalate the risk of severe illness.

Type 2 diabetes is abnormal insulin resistance which cause high glucose levels and obesity. According to CDC, there are approximately one tenth of Americans with type 2 diabetes. Diabetes is one of the risk factors which will aggravate the symptoms. Hermine et al.(2020) point out the extended duration of the disease, the increasing requirement for intensive organ support, and higher risk of mortality are the consequences of the obesity and diabetes. From the perspective of immune system, they believe that diabetes will increase the risk of more severe outcomes in COVID-19 because of “the viral toxicity on metabolically-relevant tissues including pancreatic beta cells and targets of insulin action”. Research by Guo et al.(2020) support that COVID-19 patients with only diabetes will be more likely to have acute illness like "severe pneumonia, release of tissue injury-related enzymes, excessive uncontrolled inflammation responses and hypercoagulable state associated with dysregulation of glucose metabolism. Hence, through the project, the goal is to examine the effect of the diabetes will differentiate the survival of COVID-19 patients statistically.

This project interests in how diabetes will affect the survival of the people who are infected with COVID-19 and whether among various risk factors the impact on survival brought by diabetes will be the significant. A Q&A form will be used to illustrate the effect of diabetes on survival of COVID-19 patients.

1. How will the survival probability be affected by diabetes if the events of interest is time to death?
2. Will the relative risk of the viral infection caused by diabetes change over time if the events of interest is time to death? = whether the baseline hazards are proportional for the different groups(with diabetes or w/o diabetes)?
3. Is diabetes a higher risk factor in death of COVID-19 compared to another common comorbidity, hypertension?
4. Will there exist interaction effect on the survival between diabetes and age/gender if the events of interest is time to death?
5. How long can a diabetes patient infected with COVID-19 live or recover compared to patients without diabetes?

Next, this project will perform several popular methods used in survival analysis to investigate the relationship between diabetes and survival statistically.

Methods

1. Preprocess

The original dataset(Xu et al., 2020) contains 2676311 observation among which the observation with valid follow-up time is 294. Here define the follow-up time as the time from the date of death or dishcarge to date of admission or date of confirmation. The event indicators are death and discharge. Also, the chronic comorbidities, hypertension and diabetes will be extracted from the column of chronic_disease.

2. Statistical Survival Analysis Methods

Kaplan-Meier

Kaplan-Meier estimator is a method to calculate the probability of survival(S(x)) over time based on the number at risk and number of events at each unique death time. *survfit* function can be used to implement Kaplan-Meier estamtor in R.

d_s is the number of the observations who die at time s and Y_s is the number of observations in the risk set at time s.

$$\hat{S}(x) = \prod_{s=0}^{x}(1 - P(x_i = s|x_i \geq s)) = \prod_{s=0}^{x}(1 - \frac{d_s}{Y_s})$$

```
km_diabetes = survfit(Surv(Time, d_death) ~ diabetes, data = data)
```

Log-rank test

Log-rank test is a non-parametric test to compare the differences in survival among different groups based on the ranks of the time and regardless of the time distribution. The null hypothesis might be no difference in survival among different groups or identical hazard for different groups. However, log-rank test is a directional test which implies that the alternative hypothesis is $S_1(x) = S_0(x)^\theta$ or proportional hazards.

K is the largest distinct time of observed events in either group, $O_k = (d_{1k}, d_{2k}, \ldots, d_{pk})$ and d_{ik} is the observed number of events in group i at time k. $E_{ik} = Y_{ik}\frac{O_i}{Y_i}$ where Y_{ik} is the number of subjects at risk at the start of period of k in the group i. V_k is the variance of O_k.

$$\chi^2_p = (\sum_{k=1}^{K} O_k - \sum_{k=1}^{K} E_k)'(\sum_{k=1}^{K} V_k)^{-1}(\sum_{k=1}^{K} O_k - \sum_{k=1}^{K} E_k)$$

```
logrank_diabetes = survdiff(Surv(Time, d_death) ~ diabetes, data = data)
```

Fleming-Harrington

Fleming-Harrington test is weighted log-rank tests with the weight, $w_k = \hat{S}(x_{k-1})^p(1 - \hat{S}(x_{k-1}))^q$. When q=p=0, it is equal to log-rank tests. If the p>0 and q=0, the test emphasizes early difference while if p=0 and q>0, the test emphasizes late difference. By using Fleming-Harrington test, we can check whether the alternative hypothesis of the log-rank test is valid.

Fleming-harrington test is available in *survMisc* library in R. By calling the function *comp*, the result will produce p-value for test given p and q.

```
comp(ten(Surv(T, d) ~ Z, data = data),
     p = c(0, 0, 1), q = c(0, 1, 0))
```

Cox regression

Cox model(proportional hazards regression) is a semiparametric model with an unknown baseline hazard rate which is treated nonparametrically. The ratio of two individual with covariate values Z and Z* will be $\frac{h(t|Z)}{h(t|Z*)} = \frac{h_0(t)exp(\sum_{k=1}^{p}\beta_k Z_k)}{h_0(t)exp(\sum_{k=1}^{p}\beta_k Z_k^*)} = exp[\sum_{k=1}^{p}\beta_k(Z_k - Z_k^*)]$. By using the Cox model,we can calculate the relative risk(hazard ratio).

$h(x|Z_i)$ is the hazard rate at time x for an individual with risk vector Z. $h_0(x)$ is an arbitrary baseline hazard rate.

$$h(x|Z_i) = h_0(x)exp(\beta' Z_i)$$

Cox regression model is available in library *survival* by calling the function of *coxph*.

```
cox_diabetes= coxph(Surv(Time, d_death) ~ diabetes, data = data)
```

Weibull Model

Weibull model is a parametric regression model for estimating the parameters of the conditional distribution of $x_i|Z_i$ from weibull distribution family. The accelerated failure weibull model doesn't assume proportional hazards.

X denotes the time to event and Z is avector of fixed time explanatory covariates. W has the standard extreme value distribution. $\alpha = 1/\sigma$, $\lambda = exp(-\mu/\sigma)$ and $\beta_j = -\gamma_j/\sigma$.

$$Y = lnX = \mu + \gamma^t Z + \sigma W$$

$$h_X(x|Z) = \lambda\alpha x^{\alpha-1}exp(\beta^t Z)$$

$$S_X(x|Z) = exp[-\lambda x^{\alpha}exp(\beta^t Z)]$$

Weibull regression is available in library *survival* by calling the function of *survreg* with specified dist as "weibull".

```
fit_diabetes = survreg(Surv(Time, d_death) ~ diabetes,
data = data, dist = "weibull")
```

All these method are based on the assumption of non-informative censoring that observed time and cencoring time are independent in all groups.

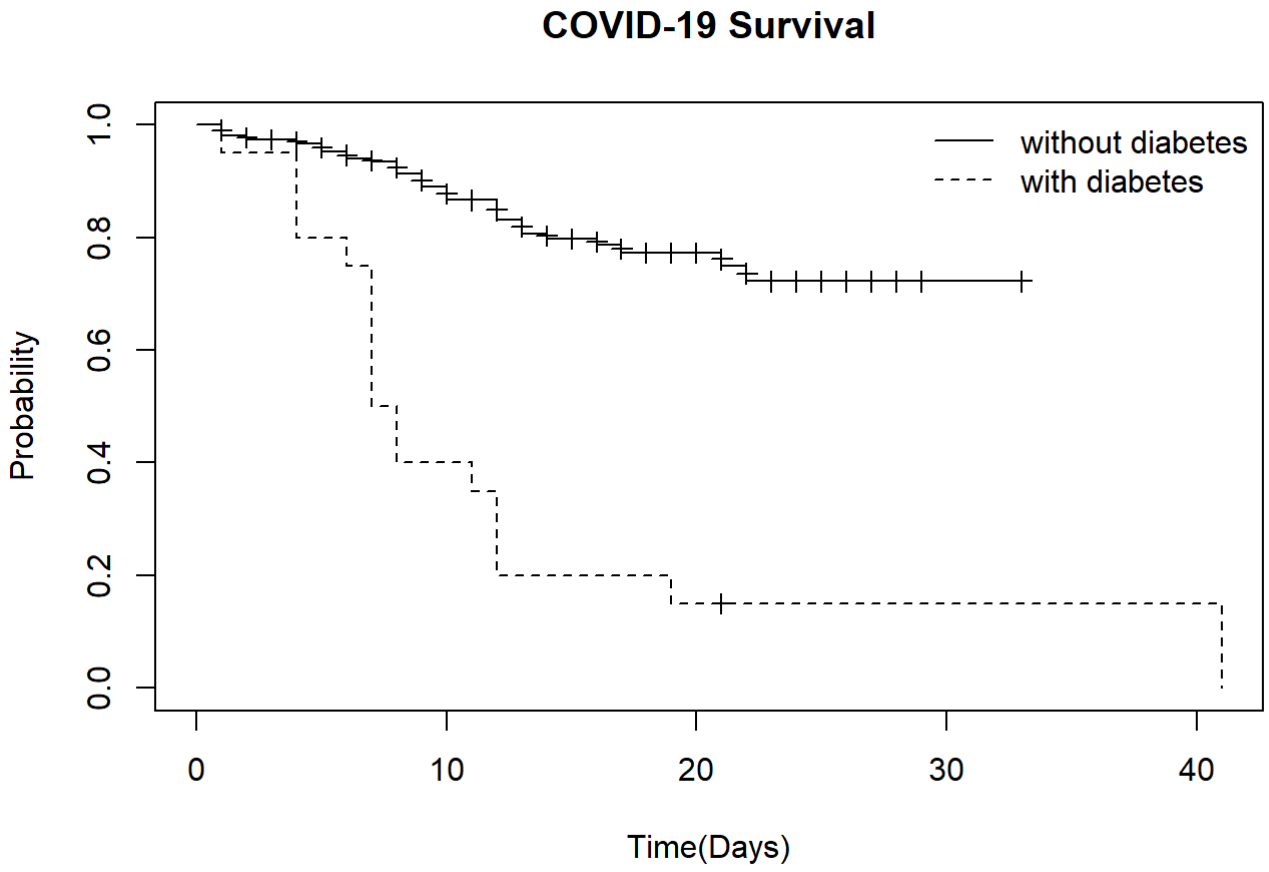
Results

1. How will the survival probability be affected by diabetes if the events of interest is time to death?

The Kaplan-Meier esitmates for the survival probability of infected observations with diabetes are much lower than the ones without diabetes. Two curves end at different points since the largest survival times of two groups are different. The figure suggests that patients without diabetes have the better favorable prognosis compared to the patients with diabetes.

```
plot(km_diabetes, lty = 1:2,mark.time = TRUE,
     main = "COVID-19 Survival",
     ylab = "Probability",
     xlab = "Time(Days)")

legend("topright", legend = c("without diabetes", "with diabetes"), lty = 1:2, bty = "n")
```



The p-value of the log-rank test is close to zero which provides significant evidenct to support the alternative hypothesis that the survival time of patients with diabetes is diffent from the patients without diabetes' survival time.

```
#p-value
pchisq(logrank_diabetes$chisq,df = 1,lower.tail = F)
```

```
## [1] 5.40121e-14
```

To check whether the log-rank test is powerful in this case statistically, the Fleming-Harrington shows the p-value for q = 0 and p = 0 is almost zero which indicates hazards ratio is constant and the alternative hypothesis for log-rank test is valid.

```
comp(ten(Surv(Time, d_death) ~ diabetes, data = data),
     p = c(0, 0, 1), q = c(0, 1, 0))
```

```
##                     Q        Var      Z      pNorm
## 1          1.3495e+01 3.2454e+00 7.4911 6.8390e-14 ***
## n          2.6260e+03 1.3665e+05 7.1038 1.2139e-12 ***
## sqrtN      1.8568e+02 6.2410e+02 7.4326 1.0658e-13 ***
## S1         1.1796e+01 2.4840e+00 7.4843 7.1942e-14 ***
## S2         1.1728e+01 2.4545e+00 7.4858 7.1054e-14 ***
## FH_p=0_q=0 1.3495e+01 3.2454e+00 7.4911 6.8390e-14 ***
## FH_p=0_q=1 1.3334e+00 6.0541e-02 5.4193 5.9849e-08 ***
## FH_p=1_q=0 1.2162e+01 2.6129e+00 7.5239 5.3069e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##                maxAbsZ        Var      Q     pSupBr
## 1          1.3612e+01 3.2454e+00 7.5559 8.3045e-14 ***
## n          2.6300e+03 1.3665e+05 7.1146 2.2446e-12 ***
## sqrtN      1.8636e+02 6.2410e+02 7.4599 1.7319e-13 ***
## S1         1.1875e+01 2.4840e+00 7.5347 9.7811e-14 ***
## S2         1.1805e+01 2.4545e+00 7.5351 9.7478e-14 ***
## FH_p=0_q=0 1.3612e+01 3.2454e+00 7.5559 8.3045e-14 ***
## FH_p=0_q=1 1.3686e+00 6.0541e-02 5.5625 5.3198e-08 ***
## FH_p=1_q=0 1.2243e+01 2.6129e+00 7.5743 7.2053e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The coefficient for the diabetes binary covariate is significant in the cox model with p-value near to zero. The hazard ratio is 6.889 which indicates that the risk of death is almost 7 times higher in patients infected COVID-19 with diabetes who might die in the study compared to those without diabetes.

```
  summary(cox_diabetes)
```

```
## Call:
## coxph(formula = Surv(Time, d_death) ~ diabetes, data = data)
##
##   n= 294, number of events= 59
##
##                coef exp(coef) se(coef)     z Pr(>|z|)
## diabetesTRUE 1.930     6.889    0.290 6.655 2.83e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## diabetesTRUE     6.889     0.1452     3.902     12.16
##
## Concordance= 0.625  (se = 0.03 )
## Likelihood ratio test= 32  on 1 df,   p=2e-08
## Wald test            = 44.29  on 1 df,   p=3e-11
## Score (logrank) test = 59.7  on 1 df,   p=1e-14
```

So based on the plot of KM estimator and log-rank test, we can conclude that the survival time for the patients with diabetes is significantly different from those without and their hazards are proportional with the hazard ratio of 6.889.

2. Will the relative risk of the viral infection caused by diabetes change over time if the events of interest is time to death? = whether the baseline hazards are proportional for the different groups(with diabetes or w/o diabetes)?

To test whether the cox regression is valid, here reformats the data and adds a time-varying covariate, $logx\ Z$. From the result, the p-value for time-varying covariate is close to zero. Hence, there is significant evidence that the null hypothesis of proportional hazards is false.

$$h(x|Z=1) = h_0(x)x^{-1.2}exp(4.7)$$

$$log(HR) = 110 - 1.2log(x)$$

```
newdata <- survSplit(Surv(Time,d_death)~.,data=data,
cut=unique(data$Time[data$d_death==1]))
newdata$diabetes_time = log(newdata$Time)*newdata$diabetes
coxph(Surv(Time, d_death) ~ diabetes+diabetes_time, data = newdata)
```

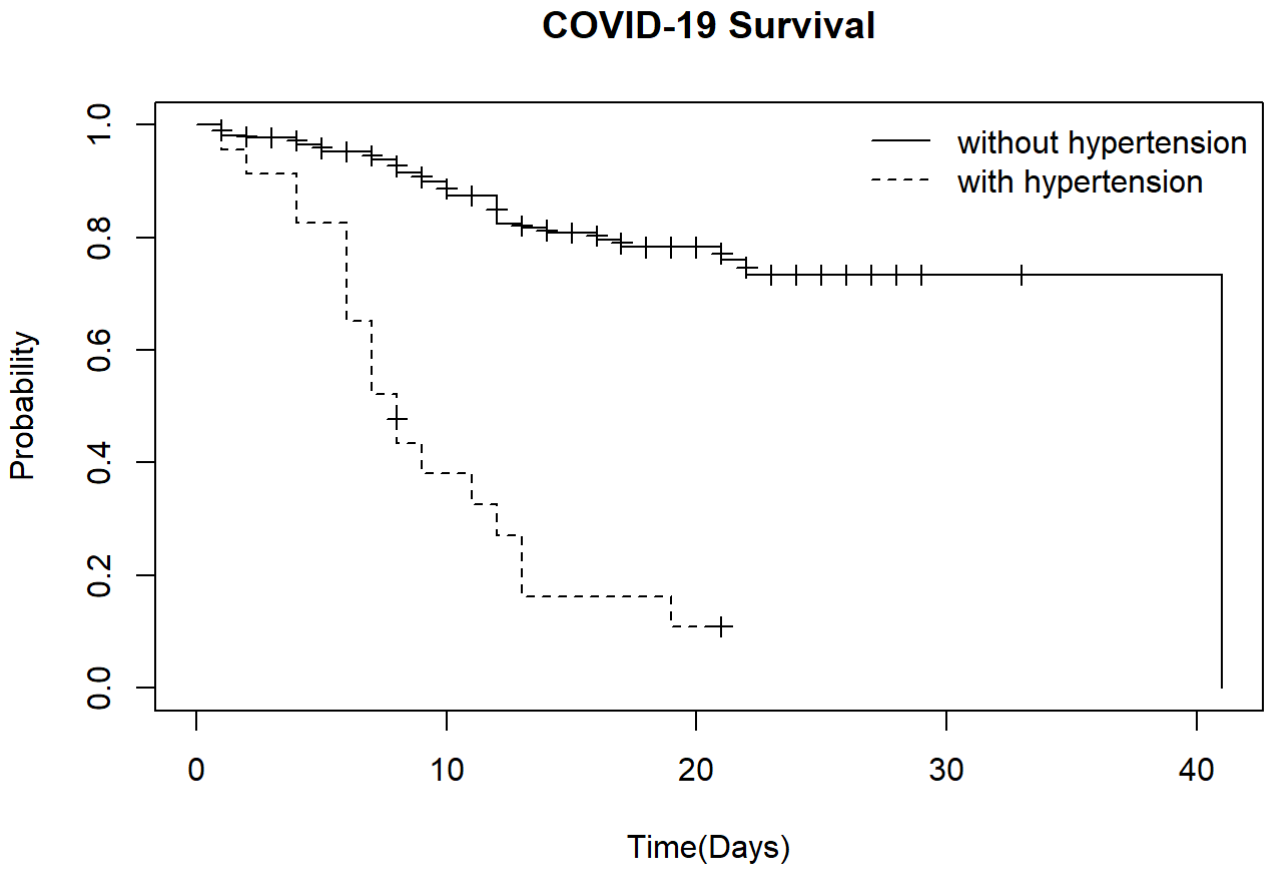
```
## Call:
## coxph(formula = Surv(Time, d_death) ~ diabetes + diabetes_time,
##     data = newdata)
##
##                   coef exp(coef) se(coef)      z        p
## diabetesTRUE    4.7031  110.2852   0.7756  6.064 1.33e-09
## diabetes_time  -1.1986    0.3016   0.3374 -3.552 0.000382
##
## Likelihood ratio test=44.82  on 2 df, p=1.847e-10
## n= 2794, number of events= 59
```

3. Is diabetes a higher risk factor in death of COVID-19 compared to another common comorbidity, hypertension?

The survival probability curve of the patients infected by COVID-19 with hypertension is far below the curve of those without hypertension.

```
km_hypertension = survfit(Surv(Time, d_death) ~ hypertension, data = data)
plot(km_hypertension, lty = 1:2,mark.time = TRUE,
     main = "COVID-19 Survival",
     ylab = "Probability",
     xlab = "Time(Days)")

legend("topright", legend = c("without hypertension", "with hypertension"), lty = 1:2, bty =
"n")
```



First, we fit the data to the weibull model with only one diabetes covariate. The coefficient of the diabetes is significant. However, when we fit the data to the weibull model with two covariates, diabetes and hypertension, the coefficient of the diabetes is unsignificant but the coefficient of the hypertension is significantly different from zero. Therefore, it suggests that hypertension has more potent effect on the survival time for patients who is infected with COVID-19 compared to diabetes.

```
summary(fit_diabetes)
```

```
##
## Call:
## survreg(formula = Surv(Time, d_death) ~ diabetes, data = data,
##     dist = "weibull")
##               Value Std. Error     z       p
## (Intercept)   3.990      0.185 21.62 < 2e-16
## diabetesTRUE -1.448      0.277 -5.23 1.7e-07
## Log(scale)   -0.225      0.105 -2.13   0.033
##
## Scale= 0.799
##
## Weibull distribution
## Loglik(model)= -279.2   Loglik(intercept only)= -294.5
##  Chisq= 30.45 on 1 degrees of freedom, p= 3.4e-08
## Number of Newton-Raphson Iterations: 8
## n= 294
```

```
fit = survreg(Surv(Time, d_death) ~ diabetes + hypertension,
data = data, dist = "weibull")
summary(fit)
```

```
##
## Call:
## survreg(formula = Surv(Time, d_death) ~ diabetes + hypertension,
##     data = data, dist = "weibull")
##                   Value Std. Error     z       p
## (Intercept)       3.965      0.179 22.17 < 2e-16
## diabetesTRUE     -0.560      0.347 -1.61 0.10702
## hypertensionTRUE -1.149      0.334 -3.45 0.00057
## Log(scale)       -0.286      0.107 -2.67 0.00763
##
## Scale= 0.751
##
## Weibull distribution
## Loglik(model)= -273.5   Loglik(intercept only)= -294.5
##  Chisq= 41.97 on 2 degrees of freedom, p= 7.7e-10
## Number of Newton-Raphson Iterations: 7
## n= 294
```

The contingency table between diabetes and hypertension shows that 70% of patients with diabetes also have hypertension while 61% of the patients with hypertension have diabetes.

```
table(data$diabetes,data$hypertension)
```

```
##
##         FALSE TRUE
##   FALSE   265    9
##   TRUE      6   14
```

4. Will there exist interaction effect on the survival between diabetes and age/gender if the events of interest is time to death?

The p-value(1.1e-06) of the Wald test show this model is a correct model compared with the model with intercept only. By adding the interaction terms, the result of the weibull model shows that only diabetes*age is significant, which means that the effect of the diabetes depends on age.

```
fit2 = survreg(Surv(Time, d_death) ~ diabetes+I(diabetes*as.numeric(age))+I(diabetes*as.numer
ic(sex=="male")),
data = data, dist = "weibull")
summary(fit2)
```

```
##
## Call:
## survreg(formula = Surv(Time, d_death) ~ diabetes + I(diabetes *
##     as.numeric(age)) + I(diabetes * as.numeric(sex == "male")),
##     data = data, dist = "weibull")
##                                        Value Std. Error     z      p
## (Intercept)                           3.8419     0.1778 21.61 <2e-16
## diabetesTRUE                          1.4831     1.2710  1.17  0.243
## I(diabetes * as.numeric(age))        -0.0393     0.0183 -2.15  0.032
## I(diabetes * as.numeric(sex == "male")) -0.1223     0.4359 -0.28  0.779
## Log(scale)                           -0.2748     0.1117 -2.46  0.014
##
## Scale= 0.76
##
## Weibull distribution
## Loglik(model)= -255.1   Loglik(intercept only)= -270.3
##  Chisq= 30.4 on 3 degrees of freedom, p= 1.1e-06
## Number of Newton-Raphson Iterations: 7
## n=263 (31 observations deleted due to missingness)
```

However, the weibull model with the main effects of diabetes and age, and the interaction effect between diabetes and age shows that age is highly significant and the effect of age doesn't depend on the diabetes.

```
fit2 = survreg(Surv(Time, d_death) ~ diabetes*as.numeric(age),
data = data, dist = "weibull")
summary(fit2)
```

```
##
## Call:
## survreg(formula = Surv(Time, d_death) ~ diabetes * as.numeric(age),
##     data = data, dist = "weibull")
##                                Value Std. Error     z       p
## (Intercept)                  6.67207    0.60230 11.08 < 2e-16
## diabetesTRUE                -1.31706    1.32320 -1.00  0.3196
## as.numeric(age)             -0.05037    0.00797 -6.32 2.6e-10
## diabetesTRUE:as.numeric(age)  0.00960    0.01799  0.53  0.5937
## Log(scale)                  -0.32004    0.11032 -2.90  0.0037
##
## Scale= 0.726
##
## Weibull distribution
## Loglik(model)= -223.1   Loglik(intercept only)= -271.9
##  Chisq= 97.53 on 3 degrees of freedom, p= 5.3e-21
## Number of Newton-Raphson Iterations: 9
## n=265 (29 observations deleted due to missingness)
```

5. How long can a diabetes patient infected with COVID-19 live or recover compared to patients without diabetes?

From the p-value(3.4e-08) of the Wald test, the weibull model with unicovariate diabetes is correct model. The weibull model shows $\hat{\mu} = 3.9898289$, $\hat{\sigma} = 0.7987409$. Given the μ and σ, $\alpha = 1.2519704$, $\lambda = 0.0067707$. Hence, for having diabetes, survival time of people infected with COVID-19 is 0.2345703 of survival time of people without diabetes.

$$Y_{death} = lnx_{death} = 4 - 1.45Z + 0.8W_{death}$$

```
fit_discharge = survreg(Surv(Time, d_discharge) ~ diabetes,
data = data, dist = "weibull")
summary(fit2)
```

```
##
## Call:
## survreg(formula = Surv(Time, d_death) ~ diabetes * as.numeric(age),
##     data = data, dist = "weibull")
##                                 Value Std. Error     z       p
## (Intercept)                   6.67207    0.60230 11.08 < 2e-16
## diabetesTRUE                 -1.31706    1.32320 -1.00  0.3196
## as.numeric(age)              -0.05037    0.00797 -6.32 2.6e-10
## diabetesTRUE:as.numeric(age)  0.00960    0.01799  0.53  0.5937
## Log(scale)                   -0.32004    0.11032 -2.90  0.0037
##
## Scale= 0.726
##
## Weibull distribution
## Loglik(model)= -223.1   Loglik(intercept only)= -271.9
##  Chisq= 97.53 on 3 degrees of freedom, p= 5.3e-21
## Number of Newton-Raphson Iterations: 9
## n=265 (29 observations deleted due to missingness)
```

Switching the event indicator to the d-discharge, the coefficient of the diabetes covariate is 0.9153. Therefore, the time to recover for COVID-19 patients with diabetes will be 2.46 of time to recover for those without diabetes.

$$Y_{recover} = ln\ x_{recover} = 2.8 + 0.9Z + 0.45W_{recover}$$

Discussion

Summary

From the previous result, we first consider the diabetes factor independently. It shows that the survival time or the hazards of the patients infected with COVID-19 between with diabetes and without diabetes are significantly different. Statistically, given the failure of the proportional hazard, the patients with diabetes are at a higher risk of dying but the risk will slowly decrease over time. To being at lower risk, subjects with diabetes are supposed to survive longer than $exp(110/1.2)$ which is considerably large. Therefore, they are always at a higher risk compared to those without diabetes. Also, the weibull regression model suggests that the survival time for patients with diabetes is 0.23 (exp(-1.45)) of survival time of the patients without diabetes and the recovery time for them is 2.46 times as much as the time that the patients without diabetes need to recover. Generally speaking, the prognosis of the COVID-19 patients with diabetes is more likely to be worse than those without a diabetes if only diabetes covariate is considered.

Next, more factors are taken into accout. Compared to other risk factors, the main effect of diabetes isn't as potent as age and another chronic disease, hypertension. Based on the previous analysis, it provides evidence to illustrate that diabetes relies on age and hypertension, since the chances for people having diabetes will increase by age while the hypertension and diabetes have some underlying causes in common and having high blood pressure is one of the risk factor for diabetes. Plus, hypertension is the most common comorbidity with COVID-19 infection followed by diabetes(Muniangi-Muhitu et al., 2020). Perhaps, effect of the diabetes is not as noticeable as the effect of other risk factors, but to further comfirm the effect of the diabetes, more observations with pure diabetes are needed.

The high risk factors affecting the survival time of the patients infected with COVID-19 are supposed to be identified in order to distribute healthcare source properly. Those who are elderly, have chronic diesease like hypertension, diabetes are more susceptible to COVID-19 and will need intense health care. This project provides the statistical insights to investigate the association of COVID-19 and diabetes for people who will continue working on the potential mechanisms behind the effect of diabetes posted on COVID-19 patients.

Trial Design

The clinical trial data using remdesivir for the treatment of COVID-19(Fig.2, John,et al. 2020) is cumulative recoveries. Therefore, to tranform it to survival function, the probability should be replaced by 1-prob. The placebo will be the data from the dataset of Xu et al. (2020). The survival times of interest now is time to discharge.

```
#type I error
alpha = 0.05
#1 - power
beta = 1-0.9
#group proportion
p = 0.5
#H0: HR = 0; HA: HR = 0.5
theta = 0.5
#accrual period
a = 5
#follow up period
f = 20

#number of deaths we need

d = ceiling((qnorm(alpha/2) + qnorm(beta))^2 / (p * (1 - p) * log(theta)^2))

remdesivir = read_csv("remdesivir.csv",col_names = FALSE)
remdesivir$X2 = 1 - remdesivir$X2

km = survfit(Surv(Time, d_discharge) ~ 1, data = data)
S_D = stepfun(km$time, c(1, pmin(km$surv, 1) ))
S_V = stepfun(remdesivir$X1, c(1, pmin(remdesivir$X2, 1) ))

S = function(x, p) {
  p * S_D(x) + (1 - p) * S_V(x)
}

prob = 1 - 1 / 6 * (S(f, p) + 4 * S(0.5 * a + f, p) + S(a + f, p))

n = ceiling(d / prob)
```

1. Will the drug lead to longer or shorter survival times?
   Since HR = 0.5 which means that the hazard of using drug group is at smaller risk than the placebo, the drug will lead to longer survival times.
2. How many deaths are needed?
   88
3. What is the total sample size that needs to be accrued?
   114

Reference

1. Xu, Bo and Gutierrez, Bernardo and Mekaru, et al(2020). Epidemiological data from the COVID-19 outbreak, real-time case information. Scientific Data, 7(106). doi.org/10.1038/s41597-020-0448-0
2. Muniangi-Muhitu H, Akalestou E, Salem V, Misra S, Oliver NS and Rutter GA (2020) COVID-19 and Diabetes: A Complex Bidirectional Relationship. Front. Endocrinol. 11:582936. doi: 10.3389/fendo.2020.582936
3. Guo W, Li M, Dong Y, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes Metab Res Rev. 2020;36:e3319. http://doi.org/10.1002/dmrr.3319 (http://doi.org/10.1002/dmrr.3319)
4. John H. Beigel, M.D., Kay M. Tomashek, M.D., M.P.H., et al. Remdesivir for the treatment of COVID-19 — final report. N Engl J Med 2020; 383:1813-1826. doi: 10.1056/NEJMoa2007764