A Simulation of Depression and Other Mental Health Diagnoses Increase Mortality Risk After Ischemic Stroke

Llasmin Lopez
Department of Statistics
University of California, Davis

Post-stroke depression occurs in approximately one third of all ischemic stroke survivors and has been linked to worse functional outcome, slower recovery, and worse quality of life. The objective is to evaluate the effect of post-stroke depression and other mental health diagnoses on survival after hospitalization for ischemic stroke in a national cohort of veterans. We test Cox PH model for the variables depression and other mental diagnoses.

Depression and Other Mental Health Diagnoses Increase Mortality Risk After Ischemic Stroke by Williams, Linda S. et all, focused on evaluating the effect of post-stroke depression and other mental health diagnoses on survival after hospitalization for ischemic stroke in a national cohort of veterans. Time origin is defined as hospital discharge after ischemic stroke and the event of interest was death. Time scale was on-study time from October 1, 1990 to September 30, 1997, the three year follow-up point.

Censoring mechanisms involved are right censoring and left truncation. Patients were considered right censored if they were alive at the three year follow up. Patients who suffered an ischemic stroke and passed before October 1, 1990, were not included in the study and so are a source of left truncation. Another variable of left truncation are patients who fall within the time scale but die within thirty days of hospitalization discharge. These patients are excluded from the study since it cannot be determined if they had pots-stroke depression, the variable of interest.

Mortality hazard ratios were modeled using Cox regression. Thus, we will use their Cox hazard ratios to recover a Cox regression model that will simulate a data set similar to the data used in the study. We will recreate the chart and image from the original paper and use the simulated data to determine if Charlson Index > 2 should be stratified against Other mental disorder or substance abuse diagnosis.

When classifying subjects, in the cases that someone had both depression and other mental disorders they were categorized solely as depressed. In the model we constructed, it is possible for a subject to have both depression and other mental disorders. It is unclear from the details provided in the paper what details were included to produce the hazard ratios for both variables. We are interested in comparing the times to event for both depression and other mental diagnoses.

Methods

Notation. As we were given Cox proportional hazard rates for all variables included in the model, we turn to the Cox PH model for left truncation and right censoring speciified by the hazard function $\lambda(t|\mathbf{Z}) = \lambda_0(t)exp(\beta^T\mathbf{Z}$, where λ_0 is an arbitrary and unspecified baseline function and $\mathbf{Z}$ denotes a set of time-independent covariates. Let T be the failure, C the censoring, and X the observed time for a subject. We assume T and C are independent given $\mathbf{Z}$. Then, the failure indicator $\gamma = I(T \geq C)$ is equal to 1 if the patient is experiences the event and 0 if they are censored.
With the simulated data, I will test for differences in time to event between the variables depression and other mental diagnosis. To do so, we will perform a hypothesis test for Cox PH model where the global null hypothesis is $\beta = \beta_0$ and the alternative global hypothesis is $\beta \neq \beta_0$. We will test our hypothesis using the Wald, likelihood ratio, and score test.
The Wald test uses the partial likelihood function to find the partial MLE $\mathbf{b}$ where $\mathbf{b}$ is of dimension p and is normally distributed with mean β and variance matrix estimated by the inverse information matrix $I^{-1}(\mathbf{b})$. The likelihood ratio test determines the difference in the -2log-likelihood: $\chi^2_{LR} = 2[LL(\mathbf{b}) - LL(\beta_0)]$. Finally, the score test simply uses the score equation and information matrix.

Simulation

A simulation study was conducted to compare the performance of both full-cohort Cox regression models. The simulation study focused on the survival outcomes generated by the Cox PH model from the paper.We assume baseline event time as $X \sim Weibull(\alpha, \lambda_0)$ for $\alpha = 1$. This assumption implies $h_0(x) = \lambda_0$. The hazard function is then specified by $\lambda(t|\mathbf{Z}) = exp(\beta^T\mathbf{Z})$ where $\beta = (-log(1.13), -log(1.13), -log(1.07), -log(1.68), -log(1.25), -log(1.08), -log(1.04), -log(1.42), -log(0.90), -log(0.68), -log(1.50), -log(1.59), -log(1.25))$ for the covariates: depression, other mental diagnoses (omd), white, CI> 2, diabetes, coronary artery disease (cad), male, hypertension, hyperlipidemia, myocardial infarct (mi),

congestive heart failure (chf), atrial fibrillation (af), and age. Age was modeled with a gamma distribution as the study had many patients around the mean age and right skewed. The remaining covariates follow a Bernoulli distribution where their probabilities are based on their proportion to the population in the cohort.
In Table 1, we see the percentages of each variable as they occur in the simulated data set. These frequencies are similar to the ones reported in the original data set.

Table 1
: Simulation Percentages of Covariates in Depression

Depression	White	CI>2	Diabetes	cad
4.6	88.87	5.9	29.35	20.52
Male	Hypert	Hyperl	mi	chf
98.19	61.35	9.25	2.36	8.62
af				
9.83				

The largest deviance from the original data set was 8% for variable white.

We were also able to calculate the Cox hazard ratios for each variable

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
dep	diagnosed depressed	1	0.17329	0.03449	25.2385	<.0001	1.189	dep diagnosed depressed
omd		1	0.15304	0.03599	18.0861	<.0001	1.165	
wht		1	0.11343	0.02499	20.6077	<.0001	1.120	
ci2		1	0.51460	0.02849	326.1664	<.0001	1.673	
diabete		1	0.22183	0.01633	184.6229	<.0001	1.248	
cad		1	0.07670	0.01865	16.9081	<.0001	1.080	
age		1	0.03837	0.0006735	3245.6799	<.0001	1.039	
male		1	0.46157	0.06880	45.0059	<.0001	1.587	
hypert		1	-0.09849	0.01559	39.9159	<.0001	0.906	
hyperl		1	-0.37285	0.02985	156.0571	<.0001	0.689	
mi		1	0.42545	0.04481	90.1359	<.0001	1.530	
chf		1	0.48120	0.02389	405.5696	<.0001	1.618	
af		1	0.25367	0.02385	113.0765	<.0001	1.289	

Figure 1. : Cox proportional hazard ratios for the simulated data

These ratios align with the ratios reported in the original paper. Similar to the paper, every variable was found significant beyond alpha level 0.0001. The last result presented from the original data is the Cumulative Hazard function stratified by depression. The following is the same plot produced by the simulated data
In Figure 2, we can see a diagnosis of depression leads to a higher cumulative hazard function than having no depression diagnosis. This feature is also present in the cumulative hazard functions from the original paper however, they are not as linear as they are here.

Simulation Results summary. Although there were some limitations to the exact construction of the Cox regression model that led to the results reported due to vague details,

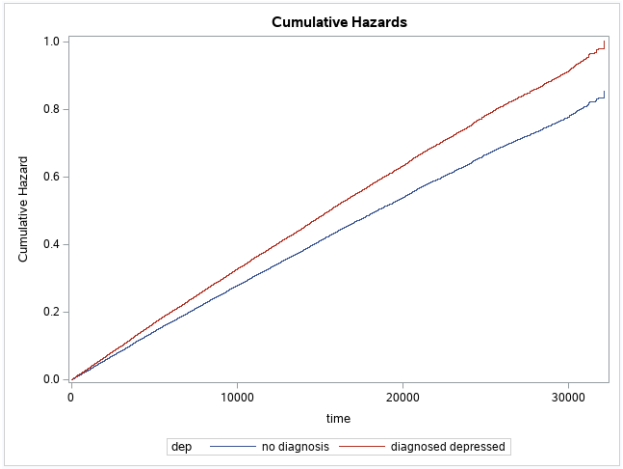

Figure 2. : Cumulative Hazards by Depression Diagnosis

the simulated data set proves to have similar characteristics as the original data based on the comparisons we were able to achieve.

Results

We compared times-to event of two varibles. We see the hypothesis test results in Figure 3.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	40.4254	2	<.0001
Score	42.5200	2	<.0001
Wald	42.4669	2	<.0001

Figure 3. : Cox proportional hazard ratios Hypothesis Test Results for Wald, Likeihood Ratio and Score

All three tests suggest, with significance greater than 99.99%, that we should reject the null hypothesis of $\beta = 0$.

Discussion

We were largely able to simulate data similar to the paper from Williams et. all. From the hypothesis tests, we see that subjects with any mental disorder diagnosis after suffering an ischemic stroke has greater hazard risk. It would be interesting to be able to stratify mental diagnoses by type of diagnoses and how many diagnoses a subject was identified for. I believe an investigation into this topic would show a dependence but unfortunately we did not have enough information to take this next step in the analysis.

Appendix:SAS

```
LIBNAME proj2; /** FOR CSV Files uploaded from Windows **/ FILENAME CSV "/home/u49996923/Project2/mydata1.csv" TERMSTR=CRLF;
/** was1 Import the CSV file. **/
PROC IMPORT DATAFILE=CSV OUT=data1 DBMS=CSV REPLACE; getnames=yes; RUN;
* check variables; proc contents data=data1;run;
* formatting variables; proc format; value depression 0="no diagnosis" 1="diagnosed depressed" ; run; proc format; value othermd 0="no omd" 1="omd" ; run; data data1; set data1; format dep depression.; format omd othermd.; run;
* age means; proc means data=data1; var age; by dep; run;
* generate frequency table; ods listing gpath='/home/u49996923/Project2/'; ods graphics / imagename="p1" imagefmt=png; proc freq data=data1; tables dep wht ci2 diabete cad male hypert hyperl mi chf af; run;
*general cox; proc phreg data=data1; class dep; model time*censor(1) = dep omd wht ci2 diabete cad age male hypert hyperl mi chf af;; run;
*multiple lines; data covs; format dep depression.; input dep omd id; datalines; 0 0 0 0 1 1 1 0 2 1 1 3 ; run;
ods listing gpath='/home/u49996923/Project2/'; ods graphics / imagename="p3" imagefmt=png; proc phreg data=data1 plots(overlay)=(cumhaz); class dep(desc); model time*censor(1) = dep; baseline covariates=covs out=base; run;
proc phreg data =data1 plots(overlay)=(survival); class dep(desc); model time*censor(1) = dep omd; baseline covariates=covs out=base /rowid=id; run;
```

R Code

```
library(MASS)
sim_cox<- function(N,lambda0, beta, censor.right)   N = Total sample size  beta = PH coefficients  lambda0 = rate parameter of the exponential distribution for baseline categorical : depression, other mental disorder (omd), white race, charlson index \geq 2, diabetes, coronary arterary disease (cad)  male,hypertension, hyperlipidemia, myocardinal infarct,congestive heart failure, atrial fibrilation
dep <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(48714/51119, 2405/51119))  0-not depressed, 1-depressed omd <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(48862/51119, 2257/51119))  0-no disorder, 1-omd wht <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(5528/51119, 45591/51119))  0-not white, 1-white ci2 <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(48156/51119, 2963/51119))  0-ci<2, 1-otherwise diabete <- sample(x=c(0,1),size=N, replace=TRUE, prob=c(36052/51119, 15067/51119))  0-no diabetes, 1-diabetes cad <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(40724/51119,10395/51119))  0-no cad, 1-cad male <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(899/51119, 50220/51119)) 0-female, 1-male hypert <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(19867/51119, 31252/51119))  0-no, 1-yes hyperl <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(46328/51119, 4791/51119))  0-no, 1-yes mi <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(49951/51119,1168/51119))  0-no, 1-yes chf <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(46777/51119,4342/51119)) 0-no, 1-yes af <- sample(x=c(0,1), size=N, replace=TRUE, prob=c(46109/51119,5010/51119)) 0-no, 1-yes
continuous covariate, mutually independent age = round(rgamma(N,shape = (67.0827769^2)/(10.0392192^2), rate = 67.0827769/(10.0392192^2)))
generate underlying event time T <- rweibull(n=N, shape=1, scale = lambda0*exp(beta[1]*dep +beta[2]*omd +beta[3]*wht +beta[4]*ci2 + beta[5]*diabete +beta[6]*cad +beta[7]*age+ beta[8]*male +beta[9]*hypert +beta[10]*hyperl + beta[11]*mi +beta[12]*chf +beta[13]*af))
censoring times ctime = runif(N, min=0, max=censor.right)
follow-up times and event indicators time <- pmin(T, ctime, censor.right) censor <- as.numeric(T>ctime | T>censor.right)
data set data.frame(id=1:N, group = dep, omd=omd, wht=wht, ci2=ci2, diabete=diabete, cad=cad, male=male, hypert=hypert, hyperl=hyperl, mi=mi, chf=chf, af=af, age=age, time=time, censor=censor)
want alpha < 1, since suspect deacreasing hazard rates, lambda > 0 bs=c(-log(1.13),-log(1.13),-log(1.07),-log(1.68),-log(1.25),-log(1.08),-log(1.04), -log(1.42),-log(0.90),-log(0.68),-log(1.50),-log(1.59),-log(1.25)) mydata=sim_cox(N=51119, lambda0=1500, beta = bs, censor.right= 32300) write.csv(mydata, file="C:/Users/Llasmin/Documents/mydata2.csv")
```